

# Synthesis Of Nucleoside Analogues: Glycosylation, Rigid Nucleosides And Janus Wedge Derivatives

Author: Ayan Pal

Persistent link: <http://hdl.handle.net/2345/2750>

This work is posted on [eScholarship@BC](#),  
Boston College University Libraries.

---

Boston College Electronic Thesis or Dissertation, 2012

Copyright is held by the author, with all rights reserved, unless otherwise noted.

Boston College

The Graduate School of Arts and Sciences

Department of Chemistry

SYNTHESIS OF NUCLEOSIDE ANALOGUES: GLYCOSYLATION, RIGID  
NUCLEOSIDES AND JANUS WEDGE DERIVATIVES

a dissertation

by

AYAN PAL

submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

June 2012

© copyright by AYAN PAL

2012

## **Abstract**

Nucleic Acids are unique biopolymers capable of encoding and transferring genetic information from one generation to the next for every form of life. This fascinating property has made them the topic of intense research from a variety of aspects. Some researchers try to understand how life might have started. Some try to elucidate how the whole process works. Some try to use the properties of nucleic acids as a tool for various purposes. The continuous effort over more than a century explored a lot about the structures and functions of nucleic acids. There is a lot to be discovered yet.

This work began with the design and development of a new class of nucleoside analogue with the goal to study their ability to bind nucleic acids. The ongoing research will establish their application as therapeutics and as biomolecular tools. Along the way significant effort went into preparing these analogues. New methodology was developed to address some of the unanswered synthetic problems of nucleoside chemistry.

## Table of Contents

Dedication and Acknowledgement.....	iv
Abbreviations and Acronyms.....	viii

### Chapter 1: Introduction

1.1 Nucleic Acids- Discovery and Brief History.....	2
1.2 Structure of DNA and RNA.....	5
1.3 Different forms of DNA.....	8
1.4 Nucleoside Analogues.....	16
1.5 Peptide Nucleic Acid.....	22
1.6 Factors Determining the Modified Nucleic Acids Binding Properties.....	30
1.7. Potential Applications of Modified Oligonucleotides as Therapeutics.....	32
1.8. Application of Modified Oligonucleotides as Diagnostic Tool.....	34
1.9. Applications of PNA.....	35
1.10. Specific Examples of Modified Nucleosides Which Have Medicinal Applications.....	39
1.11. Scope of the Research.....	41

### Chapter 2: Bicyclo-1,1'-dicarboxylic Acid Cyclicanhydride as a Directing/Protecting

#### Group Purine Glycosylations

2.0. Background.....	54
----------------------	----

2.1. Common Routes to prepare 2'-deoxyribonucleosides.....	56
2.2. Problems with Traditional Routes for Modified Nucleoside Synthesis.....	60
2.3. Development of Bulky Substituent at C6-position of Purines.....	61
2.4. Screening New Bulky Substituent at C6-position of Purines.....	64
2.5. Bicyclo-1,1'-dicarboxylic Acid Cyclicanhydride as a Directing/Protecting Group.....	70
2.6. Experimental.....	77

### **Chapter 3: Cyclic Oligonucleotides**

3.1. Rigidified Nucleic Acids.....	108
3.2 Synthesis of 8,5-( <i>S</i> )-cyclo-2'-deoxyguanosine.....	113
3.3 Synthesis of 6,8'-cyclo-2'-deoxyadenosine.....	118
3.4 Experimental and Methods.....	122

### **Chapter 4: Purine-like Janus Wedge Residues**

4.1 Triplex Forming Oligonucleotides.....	152
4.2 Janus Wedge (Jw) Triplexes.....	153
4.3 Limitation with Pyrimidine-like Jw Triplexes.....	159
4.4 Preparation of Monomer to target T-A base pair Y1.....	161

<b>4.5</b>	<b>PNA Synthesis incorporating Y1.....</b>	<b>165</b>
<b>4.6</b>	<b>Preparation of Modified Base to target A-T base pair.....</b>	<b>168</b>
<b>4.7.</b>	<b>Preparation of Modified Base to target C-G base pair.....</b>	<b>174</b>
<b>4.8.</b>	<b>Attempt to Synthesize Modified Base to target G-C base pair.....</b>	<b>176</b>
<b>4.9.</b>	<b>Experimental and Methods.....</b>	<b>179</b>

*Dedicated to My Family and Teachers*



## **Acknowledgement**

The journey of the graduate life has been incredible. It had the moments of highs and lows. And it is always great to find someone around the corner. I wonder how everyone you meet or work influences your life and makes you the person who you are today. I am glad to be at this stage of my life and grateful to every one of them.

It is impossible to be writing today without paying my regards to Prof. McLaughlin. I would like to thank him for giving me the opportunity to work in his laboratory. He is truly a guide; a person who showed me the path to become a scientist and a person I can always come to whenever I need advice. Needless to say his guidance has been invaluable to the work I will be presenting in my thesis.

I would like to thank Profs. McLaughlin, Roberts, Gao and Christianson for being in my Graduate Committee. Prof. Gao read my progress report every year and gave me his thoughts. Prof. Weerapana was very helpful. She analyzed a few sample for one of my projects during a weekend. Profs. Gao and Weerapana always were very supportive whenever I dropped in their office for their advice. Prof. Roberts is always encouraging.

I have had the opportunity to work with a number of talented post-docs, grad students and undergrads. It is incredible that I can reach any of our past groups members any time I needed advice or anything. I would to thank Dr. Meena, Dr. Han Chen, Dr. Greco, Dr. Horhota and Dr. Arico for their support. I cannot thank enough to Dr. Christianson who has helped me a lot during last stage of my graduate school. He has been doing an incredible job managing our lab and going through the manuscripts and thesis. All these wonderful people make your life so easy.

I am also lucky to have Han, Eric, Chris, Kerry and Hongchuan as labmates for a long stretch of time. They are not only talented chemists but also fun and encouraging. Later Nick, Alena and Yiran joined our lab and doing exceedingly well. I also need to mention about Rabih and Kenny, the two undergrads whom I worked with. They made teaching chemistry a lot of fun.

Research would have been impossible without the support from our instrumentation departments. I would like to thank Dr. Boylan, our Director of NMR Center, Mr. Domin, Director of Mass Spectrometry and Dr. Li, Director, Center for X-Ray Crystallography for their continuous support and advice.

As an international student it is always difficult to first come to a new country and find your way in. I heard a lot of these stories before I came to here. But I have to say I never felt that way. There were cultural differences. But the Department of Chemistry at Boston College made sure I have to feel that way. Everyone, from Professors to Administrative personals to International Student Center, made my life as easier as anyone can imagine. The warmth of my classmates made my initial days so much enjoyable.

My quest for Science in general started when I was in my high school. I remember Ms. G. Uma, our chemistry teacher, influenced me the most. She used to spend hours after the school to make sure that every one learns their lesson. I also like to thank my Professors at my undergraduate school, Jadavpur University, and my masters' school, IIT Kanpur. They are the people who encouraged me to become a chemist. I would like to express my gratitude to Prof. Bharadwaj, Prof. Murthy, Prof. Gupta-Bhaya and Prof. Khan for their teachings, support and guidance.

Staying in Boston can be challenging sometimes. And it is great to know that whenever you are overwhelmed for any reason there are a few people you can always count on. Joby is always

there guiding me whenever I feel low or I had any issues regarding immigration. My three years long roommate Ritaban and two years long roommate Aritra became my best friends here in Boston. They made staying in an apartment a lot of fun. During the last year or so Ranjan and Erika became very close friends of me. Hanging out with them is always stress-relieving.

There are always a few people who you really can never thank because they define your existence. They are your family. It is impossible to acknowledge their importance in a sentence or two. Without their unconditional love and support I could not have accomplished anything. My parents have always been beside me and made me the person I am today. Life would not have been enjoyable and cheerful without my wife, Samragnee. Even though she had to stay in India for most of the time, she made sure that I never felt alone. Her presence inspires me to go for the unknown.

Boston

Ayan Pal

16<sup>th</sup> May, 2012.

## Abbreviations and Acronyms

A	adenosine
Ac	acetyl
ACN	acetonitrile
Ad	adenine
AIBN	azobis(isobutyronitrile)
Bn	benzyl
Boc	<i>tert</i> -butyloxycarbonyl
Bu <sup>t</sup>	tert-butyl
Bz	benzoyl
calcd	calculated
Cbz	carbobenzyloxy
CD	circular dichromism
CDCl <sub>3</sub>	chloroform-d <sub>6</sub>
Cy <sub>2</sub> SA	bicyclohexyl-1,1'-dicarboxylic acid cyclicanhydride
Cy <sub>2</sub> SI	icyclohexyl-1,1'-dicarboxylic acid cyclicdiimide
d	day(s)
dA	2'-deoxyadenosine
dC	2'-deoxycytidine
dG	2'-deoxyguanosine
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
DCM	dichloromethane
DIPEA	diisopropylethylamine

DMAP	4-(dimethylamino)pyridine
DMF	<i>N,N</i> -dimethylformamide
DMTr	4,4'-dimethoxytrityl
DMSO	dimethylsulfoxide
DNA	$\beta$ -D-2'-deoxyribonucleic acid
dNTP	2'-deoxyribonucleoside 5'-triphosphate
DPC	diphenylcarbamoyl
Et	ethyl
Et <sub>2</sub> O	diethyl ether
EtOAc	ethyl acetate
Fmoc	Fluorenylmethyloxycarbonyl
G	guanosine
hr	hour(s)
Hex	hexanes
HPLC	high performance liquid chromatography
HRMS	high-resolution mass spectrum
M <sub>4</sub> SA	2,2,3,3-tetramethylsuccinic anhydride
M <sub>4</sub> SI	2,2,3,3-tetramethylsuccinimide
M	molar
Me	methyl
min	minute(s)
MeOH	methanol
NMR	nuclear magnetic resonance

NBS	<i>N</i> -bromosuccinimide
Pht	phthaloyl
Pu	purine
Py	pyrimidine
pyr	pyridine
$R_f$	retention factor
rt	room temperature
T (or dT)	thymidine
TBAF	<i>n</i> -tetrabutylammonium fluoride
TBS, TBDMS	<i>tert</i> -butyldimethylsilyl
TEA	triethylamine
Tf	triflate (trifluoromethanesulfonyl)
TFA	trifluoroacetic acid
TLC	thin layer chromatography
THF	tetrahydrofuran
TMS	trimethylsilyl, also tetramethylsilane
Tol	<i>para</i> -toluoyl
Ts	tosyl
RNA	$\beta$ -D-ribonucleic acid
U	uridine
UV	Ultraviolet

# **Chapter 1**

## **Introduction**

## **1.1. Nucleic Acids- Discovery and Brief History**

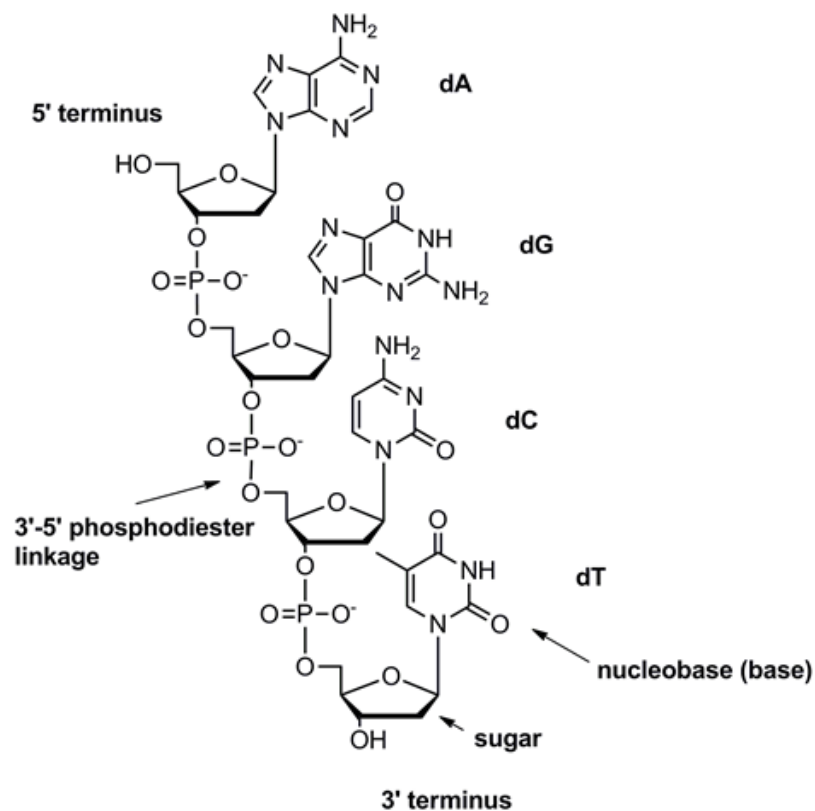
Nucleic acids are considered to be the foundation of every life form on earth. 2-deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) constitute the family, nucleic acids. In its simplest state, DNA is an ensemble of nucleic acids, water and salts, and the structure of the ensemble is dynamically dependent on all three components, e.g., sequence of nucleic acid bases, type and concentration of salts and the level of hydration<sup>1</sup>. DNA is a polynucleotide containing the genetic instructions used in the development, functioning and replication of all known living organisms (with the exception of RNA viruses). Genes are the DNA segments that carry this genetic information. Other DNA sequences have structural purposes, or are involved in regulating the use of this genetic information. In the modern era, nucleic acids have become an exciting area of research for its diverse applicability in a wide spectrum of fields<sup>2</sup> such as medicine, biology, chemistry, biotechnology, molecular biology, nanotechnology, material science etc.

In 1869 DNA was first isolated by the Swiss physician Friedrich Miescher. The isolation of a phosphorus and nitrogen containing acidic substance that was common to many discrete cell types began the study of the substance he named "nuclein"<sup>3</sup>. Albrecht Kossel isolated the non-protein component of nucleic acid and later isolated its five primary nucleobases in 1878<sup>4</sup>. The use of dyes capable of staining nuclein, as well as the rod shaped structures located in the nucleus of cells during certain stages of cell division led Flemming to conclude that "inheritance may, perhaps, be affected by the physical transmission of a particular compound from parent to offspring". These early discoveries



led others to continue their investigations into this substance, which eventually shed light on both the structures and functions of nucleic acid polymers.

Research into nuclein remained steady during the first half of the 19th century. After four decades, in 1919, Phoebus Levene identified the base, sugar and phosphate nucleotide unit<sup>5</sup>. He suggested that DNA consisted of a string of nucleotide units linked together through the phosphate groups, and resulted in the discovery of pentose carbohydrates linked through phosphodiester bonds (Figure 1.1 ) forming long polymers capable of weighing thousands of

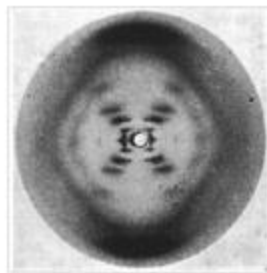


**Figure 1.1.** Single DNA strand illustrating the three major components that make up an oligonucleotide- nucleobase, carbohydrate, and phosphodiester linkage along the backbone. The phosphates link the nucleosides together through 3'-5' linkages for both DNA and RNA.

Daltons. In 1927 Nikolai Koltsov proposed that traits would be inherited via a "giant hereditary molecule" which would be made up of "two mirror strands that would replicate in a semi-conservative fashion using each strand as a template"<sup>6</sup>. Next year, Frederick Griffith discovered that traits of the "smooth" form of the *Pneumococcus* could be transferred to the "rough" form of the same bacteria by mixing killed "smooth" bacteria with the live "rough" form<sup>7</sup>. DNA's role in heredity was confirmed in 1952, when Alfred Hershey and Martha Chase showed that DNA is the genetic material of the T2 phage<sup>8</sup>.

Chargaff<sup>9</sup> discovered a consistent proportionality between adenine (A) and thymine (T) as well as cytosine (C) and guanine (G) that supported Gullard's earlier belief in hydrogen bonding interactions between purines (A and G) and pyrimidines (C and T). The miss-assignment of the thymine and guanine bases in their enol rather than keto form prevented the breakthrough that would soon revolutionize the study of nucleic acids and biochemistry.

To explain a lot of phenomenon it was necessary to confirm the structure of DNA by X-ray crystallography. In 1937, William Astbury produced the first X-ray diffraction patterns that showed that DNA had a regular structure<sup>10</sup>. After about 15 years, James D. Watson and Francis Crick suggested what is now accepted as the first correct double-helix model of DNA structure in 1953. Their key piece of evidence came from a crystal structure obtained from their collaborators Wilkins and Franklin (Figure 1.2)<sup>11</sup>. The X-ray diffraction pattern implied a helix with two periodicities along the major axis, a major one with a length of 3.4 Å and a secondary one of 34 Å.

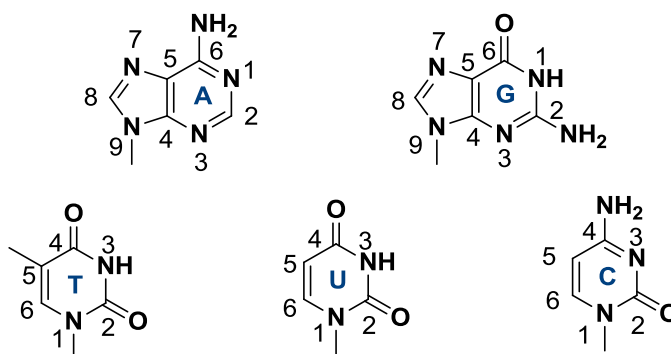


**Figure 1.2.** Photograph of the X-ray diffraction pattern observed by Wilkins and Franklin.

## 1.2. Structure of DNA and RNA

Structurally the building blocks of single strand DNA (Figure 1.1) and RNA are nucleotides. Nucleotides have three major components - nucleobase, sugar moiety, and the phosphodiester linkage. Nucleobase tethered to the sugar is called nucleoside. Nucleosides with a phosphate at C5' are the unit of a nucleic acid polymer- nucleotide. The difference between DNA and RNA lies in the fact that the latter has a 2'-hydroxyl in the sugar moiety instead of hydrogen.

### *Nucleobases*

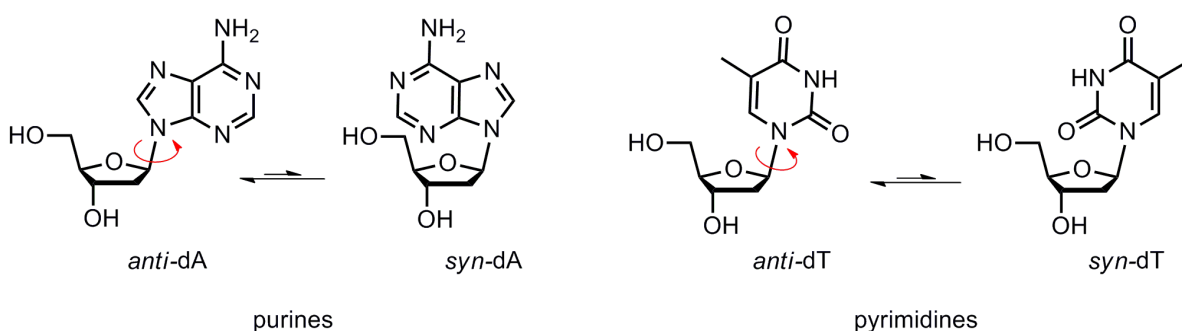


**Figure 1.3.** Structure and conventional numbering nomenclature of nucleobases (Top- Purine: adenine and guanine respectively; Bottom- Pyrimidine: thymine, uracil and cytosine respectively).

There are two classes of natural nucleobases- single ringed pyrimidines and double ringed purines. Thymine, cytosine and uracil constitute the pyrimidines whereas adenine and guanine constitutes the purines. The structure and the conventional numbering pattern have been summarized in Figure 1.3. In DNA, all these nucleobases exist except uracil. On the other hand in RNA, all the nucleobases exist except thymine.

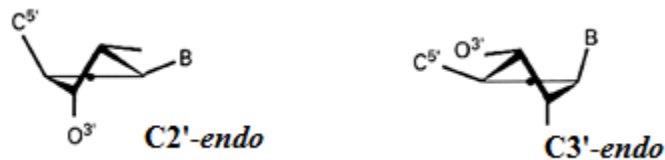
### ***Sugar or the Carbohydrate Moiety***

The ribose portion of nucleic acids (DNA and RNA) forms the scaffold that connects the nucleobases to the phosphodiester backbone linking the polymer together. The sugar unit for the RNA is the D-ribofuranosyl sugar whereas the DNA lacks the 2' hydroxyl of that sugar. Nucleobases are coupled to the sugar through glycosidic bonds between the C1' position on the sugar and the N9 or N1 positions of purines and pyrimidines respectively. The nucleobases can rotate about the glycosidic bonds; however, they are predominately found in two major regions called *syn*- and *anti*-positions (Figures 1.4), with the *anti*-position being favorable due to reduced steric clash between the sugar and nucleobase. Nevertheless, examples of both these conformations are known in nature.



**Figure 1.4.** Rotation about the glycosidic bond. The *anti*-conformation allows the standard Watson-Crick hydrogen bonding whereas *syn*-conformation can be favored under certain conditions, like a bulky substituent at C8 of purine or C5 of pyrimidines or under high salt concentration or lower hydration.

The sugar or the carbohydrate moiety can exist in a number of different conformations. The two most common conformations or sugar pucker are C2'-*endo* and C3'-*endo* (Figure 1.5). The sugar conformation adopted in B-form DNA, the most common form, and RNA are C2'-*endo* and C3'-*endo* respectively.



**Figure 1.5.** Most common sugar pucker- C2'-*endo* and C3'-*endo*.

### ***Phosphate Backbone***

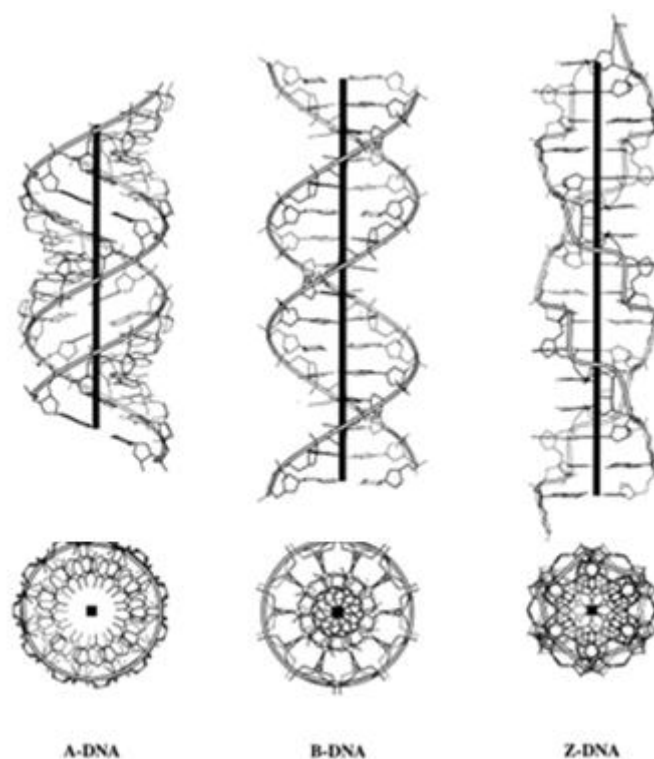
The final component of a nucleic acid polymer is the phosphodiester linkage connecting the sugar moieties and forming the biopolymer. (Figure 1.1) These negatively charged phosphates<sup>12</sup> impart a chemical stability toward hydrolysis that is thought to be critical for genome stability.

It is important to emphasize that there are many advantages of having a phosphate group as a linker rather than other alternatives like esters, amides, citric esters, arsenic esters, silicic esters etc. There is a wide range of reasons why these alternatives have limitations<sup>13</sup>. The limitations include, not limited to, higher hydrolytic rates, being neutral, toxicity, specificity issues etc. The existence of a genetic material such as DNA requires a scaffold for a connecting link that is at least divalent. It needs to be charged in order to be confined within the membrane, and therefore the linking unit should have a third, ionizable group. In order to make it hydrolytically stable the charge should be negative and should be physically close to the ester groups. All of these conditions are

met by phosphoric acid, and no suitable alternative is obvious. Another reason why nature prefers phosphate esters is its unique balance of reactivity and stability. Nevertheless, a biochemical system must not be so stable that it cannot be taken apart. DNA must be metabolized. Understanding the mechanisms of enzyme action is among the important challenges for physical organic chemistry today, but the fact that enzymes exist that can hydrolyze the diesters of phosphoric acid is sufficient to suggest that it is kinetically stable and thermodynamically labile.

### 1.3. Different forms of DNA

The three most common forms of DNA are below in (Figure 1.5): A-form, B-form, and Z-form. Under physiological conditions, the B-form is the most prevalent form of DNA.

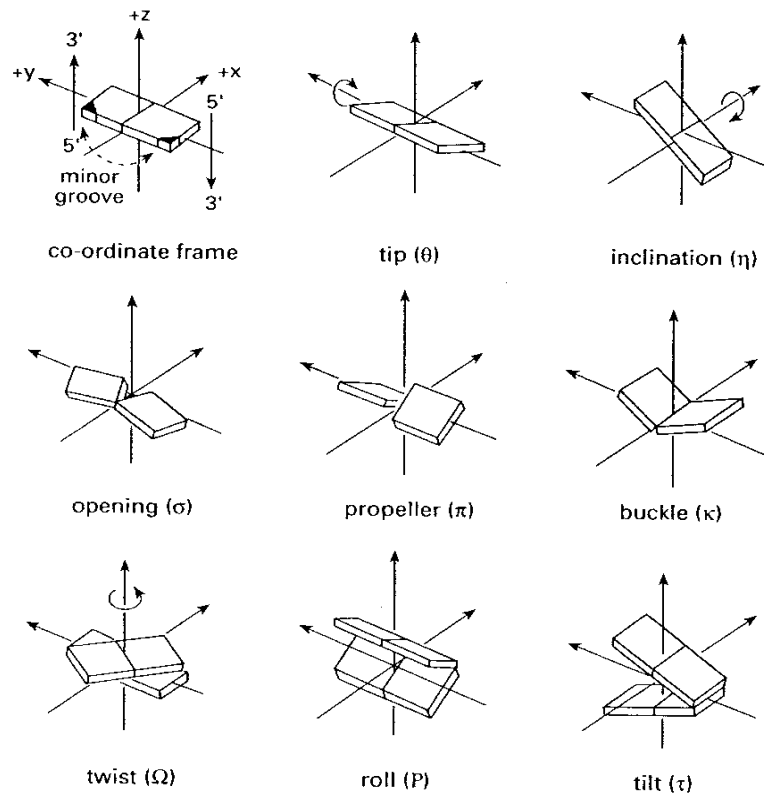


**Figure 1.5.** Structures of of A-DNA, B-DNA and Z-DNA (left to right)<sup>14</sup>. Top: Side view, Botton: Top view.

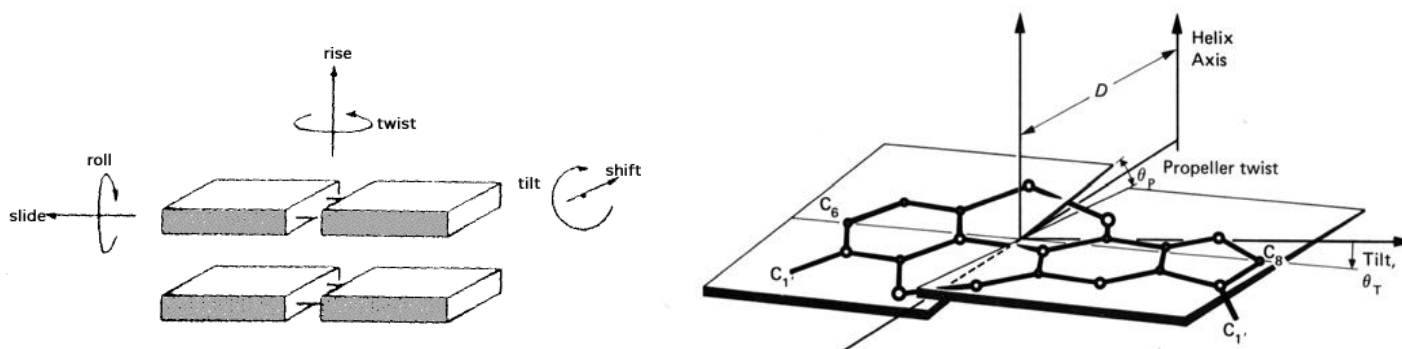
Before describing the features of these helices, it is important to introduce a few parameters that are important.

### ***Structural Parameters***

Base *tilt* is rotation about the short axis through the base pair. *Roll* is the rotation of the base pair about the long axis normal to the helical axis. *Propeller twist* is the rotation of one base with respect to the other in the same base pair. *Rise* is the distance between each base pair along the helical axis. Helical *twist* is the rotation around the helical axis. Helical *pitch* is the distance per helical repeat. *Slide* and *shift* are displacements of the base pair in the plane normal to the helical axis (Figures 1.6. and 1.7.).



**Figure 1.6.** Some of the structural parameters used with nucleic acids.



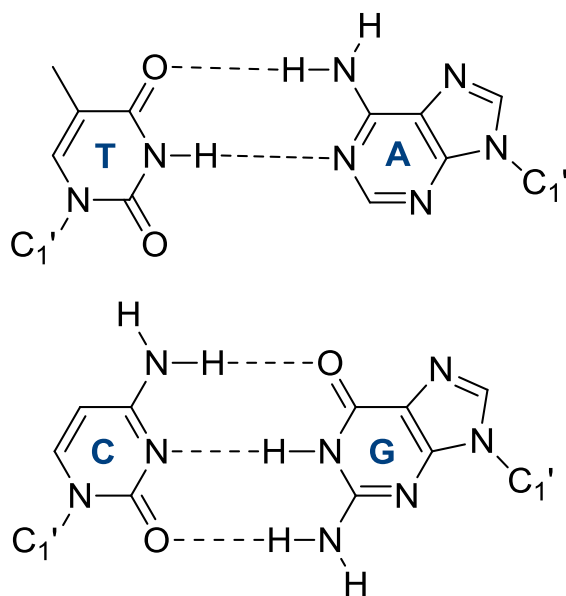
**Figure 1.7.** Left: Illustration of Rise, Roll, Slide Twist, Tilt and Shift. Right: Closer look at Propeller twist and Tilt.

### ***Sugar Pucker***

It has been discussed in section 1.2.

### ***Base Pair (bp)***

The two base pairs and the hydrogen bond involved in the pairs (Figure 1.8).

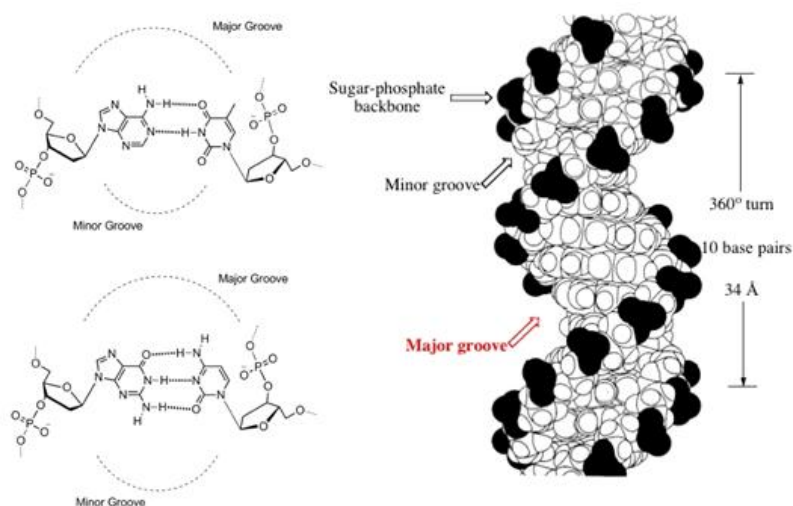


**Figure 1.8.** Two base pairs observed in DNA duplex.



## Grooves

As the two strands of DNA duplex are not directly opposite each other, the grooves are unequally sized. The major groove occurs where the backbones are far apart, the minor groove occurs where they are close together (Figure 1.9). The grooves twist around the molecule on opposite sides. The narrowness of the minor groove means that the edges of the bases are more accessible in the major groove. As a result, proteins can bind to specific sequences in double-stranded DNA usually by making contacts to the sides of the bases exposed in the major groove<sup>15</sup>.



**Figure 1.9.** Illustration of Major and Minor grooves in a DNA duplex.

In all forms the base pairing of nucleotides requires that the bases be nearly perpendicular to the major axis of the duplex, and each base pair must be nearly linear and coplanar. With these parameters it is possible to describe the structural features and differences between A, B, and Z-form DNA.

B-form is the most common structure of DNA duplex and it is right-handed helix. B-form DNA is narrower and elongated than A-form due to a C2'-*endo* sugar puckering placing

the backbone hydroxyls 7.0 Å apart. This form of helical DNA places the base pairs at the center of the vertical axis, resulting in more equally sized major and minor grooves. The base pairs are nearly perpendicular to the vertical axis of the helix with a small inclination of  $-1.2^\circ$ . B-form DNA is less distorted and more symmetrical than both A- or Z-form helices. Some of the other features have been summarized in Table 1.1.

A-form is a right-handed helix with the widest diameter and smallest rise per base pair of any form, both attributed to the sugar pucker. The C3'-*endo* sugar pucker places the 3'- and 5'- hydroxyls within 5.9 Å of one another, leading to its small rise and wide diameter. The large diameter places the base pairs off-center of the vertical axis, resulting in exaggerated major and diminished minor grooves. The major groove is very deep, while the minor groove is extremely shallow. Base pairs in an A-form conformation are on a severe incline of  $+19^\circ$ . Other important features and a comparison between the different form of DNA have been summarized in Table 1.1.

Z-form has remarkable differences compared to the other two forms. They are the most distorted form of DNA within cellular conditions. Although it is thought that it relieves supercoiling of DNA while transcription occurs<sup>16</sup>, but the actual role in nature is still unclear. Z-DNA has left-handed helicity which is generally disfavored. Alternating poly(dGC) sequences in the presence of high salt can encourage B- to Z-form transitions. Interestingly, the sugar pucker in these alternating poly(dGC) sequences is C2'-*endo* for dC residues and C2'-*exo* for dG residues. Z-form DNA is the only naturally occurring helix with *syn* conformation of dG residues. These features cause the repeating unit of Z-form to be 2 base pairs, as opposed to a repeating unit of 1 base pair for A- and B-forms. It has the narrowest diameter of all 3 forms and a severe degree of rotation per base pair

(60° versus ~35° for A- and B- forms). The Z-DNA conformation has been difficult to study because it does not exist as a stable feature of the double helix. Instead, it is a transient structure that can be induced by binding with certain proteins that reverts back to B-form afterwards<sup>17</sup>.

**Table 1.1.** Features of A, B and Z-DNA<sup>18</sup>.

	<b>A</b>	<b>B</b>	<b>Z</b>
<b>Helicity</b>	Right-handed	Right-handed	Left-handed
<b>Repeating Unit</b>	1 bp	1 bp	2 bp
<b>Rotation/bp</b>	33.6°	35.9°	60°/2
<b>Average bp/turn</b>	10.7	10.5	12
<b>Conformation about Glycosidic Bond</b>	Anti	Anti	C: anti, G: syn
<b>Sugar pucker</b>	C3'- <i>endo</i>	C2'- <i>endo</i>	C: C2'- <i>endo</i> , G: C2'- <i>exo</i>
<b>Diameter</b>	26Å	20Å	18Å
<b>Rise/bp</b>	2.3 Å	3.4 Å	3.8 Å
<b>Pitch</b>	24.6 Å	35.7 Å	45.6 Å
<b>Tilt</b>	19°	-1.2°	-9°
<b>Propeller twist</b>	18°	16°	0°

### ***Factors Determining the Structure of DNA Duplex***

There are a lot of factors involved to determine the overall dynamically favored structure of DNA duplex. It is to be emphasized here that the forms of DNA are not static structures. It is, rather, an equilibrium of various states of dynamic structures. The major factors that are responsible include base-pairing,  $\pi$ -stacking, steric effects, charge-charge repulsion, hydration, salt concentration, super-coiling, metal binding, temperature, interaction with external entities like proteins, the DNA sequence etc. The issues get even more complicated because DNA can bend, compress, stretch, and unwind differently under the influence of these factors.

**Base-pairing**<sup>19</sup>: Base-pair interaction is in the plane of the bases (horizontal) due to hydrogen bonding. Adenosine base pairs with thymidine through two hydrogen bonds, whereas guanosine base pairs with cytosine using three hydrogen bonds (Figure 1.3.3.1.). As a proof that this actually contributes to the stabilization of DNA duplex, it has been observed in numerous instances that sequences rich in G-C base pairs have been shown to be more stable than sequences rich in A-T base pairs because of the presence of a third hydrogen bond. Another indication that hydrogen bonding affects duplex stability is that longer sequences of polynucleotides are more stable than shorter sequences due to the presence of additional hydrogen bonds.

**Base stacking**<sup>19</sup>: Base stacking interactions are perpendicular to base planes stabilized mainly by  $\pi$ - $\pi$  interaction, London dispersion forces and hydrophobic effects. At its core of DNA, there is a series of aromatic rings (the nucleobases) stacked on each other. It is believed that the  $\pi$ - $\pi$  interactions results from interactions of the electron cloud above and below the plane of the molecule. These effects give a net effect that stabilizes the duplex and keeps the structure intact. The hydrophobic effect certainly plays a role in the aggregation and stacking of oligonucleotide heterocycles in addition to the contributions from  $\pi$ - $\pi$  interactions.

**Charge-charge repulsion**: The backbone of the DNA is a negatively charged phosphodiester linkage. The proximity of negatively charged backbones results in charge-charge repulsion. The concentration of cations within the environment is directly proportional to the stability of the duplex. Increasing the concentration of salt within solution has been shown to give higher melting temperatures<sup>20</sup>. These cations have been

shown to closely associate with the negative backbone, as well as in regions where the two strands are within their closest proximities<sup>21</sup>.

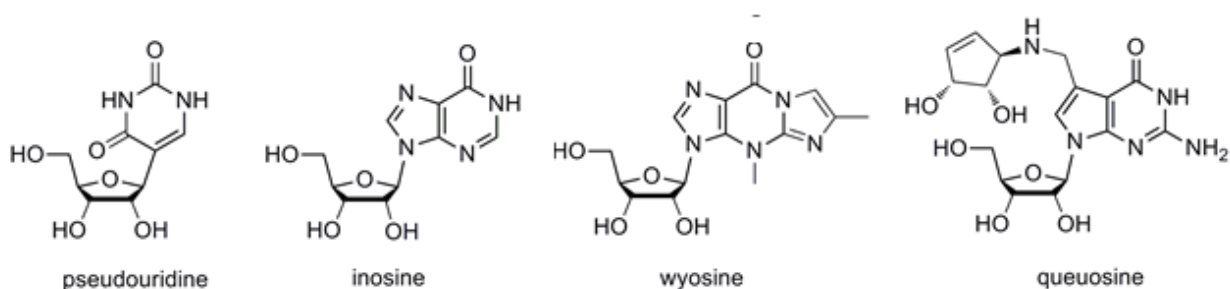
***DNA Hydration:*** Hydration of DNA is a special feature of the minor grooves. It is an important factor in determining the duplex structure (A-form or B-form) and contributes to the overall stability of the duplex by adding additional hydrogen bonding interactions with exposed functional groups on the exterior of the duplex. It is especially important for the stabilization of minor groove of A-tract DNA. The minor groove of A-tract DNA provides a unique chemical environment. The polarity and electronegative potential are high. The floor and walls of the groove are lined with hydrogen bond acceptors but are devoid of hydrogen bond donors<sup>22</sup>. A stabilizing two-tier ‘spine of hydration’ was postulated from medium and low resolution X-ray diffraction data<sup>23</sup>. The primary shell or spine of hydration is defined as the nearest layer to the duplex engaged in direct interactions with functional groups on the surface of the duplex and often with other waters in the primary shell. The secondary shell is the next layer of water engaged with the first, and so on.

The supercoiling and bending of DNA becomes vitally important in the cellular processing and storage of its genetic material. Moreover, a great deal of packaging and compression is necessary. DNA is wound tightly around spool-like proteins known as histones, which are then further packaged as chromatin and compacted into chromosomes.

## 1.4. Nucleoside Analogues

A plethora of examples of nucleosides are known that differ to various extents from the common canonical nucleosides. These are referred to as nucleoside analogues. The variation may be simple as a replacement of an atom by another, for example replacing the sugar ring oxygen by carbon or sulfur, or change in the bond order, for example connecting the C2' and C3' by a double bond. The modification can be fairly complicated, for example connecting the C2'-oxygen and the C4' by a methylene group of the ribonucleoside or extending the aromaticity of nucleobases or tethering different functionally important groups at specific positions of the nucleobases. Some of the examples have been discussed in this section.

Some of examples of these analogues exist naturally in tRNA (Figure 1.10). Inspired from nature, chemists have also developed wide range of nucleoside analogues altering the nucleobase, sugar moiety, the backbone or a combination of all three of them.



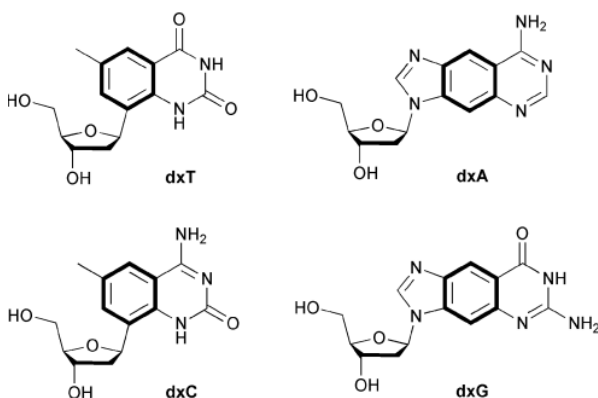
**Figure 1.10.** Structures of a few nucleoside analogues present in t-RNA.

### 1.4.1. Nucleobase Modification

In order to study the effect of modified nucleobases on Watson-Crick(W-C) base pairing and in part, to expand the genetic code, Kool and co-workers synthesized a range of

expanded nucleobases (Figures 1.11 and 1.12)<sup>24</sup>. They designated the three classes of these nucleobases: xDNA (expanded DNA)<sup>25</sup>, yDNA(wide DNA)<sup>26</sup> and yyDNA( double wide DNA)<sup>27</sup>.

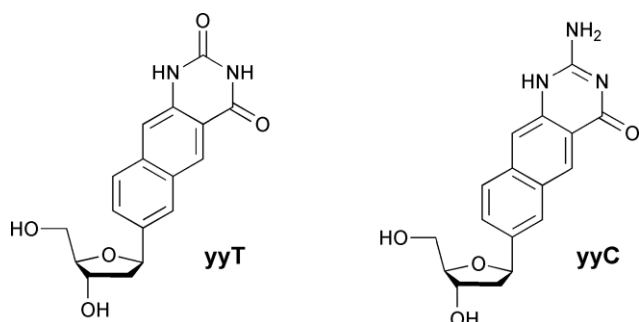
They found that self-complementary xDNA 10-mers, in which each of the nucleobases was ‘horizontally’ expanded by the width of a phenyl ring (2.4 Å), still formed a standard B-DNA duplex. In fact, the helical right-handedness, deoxysugar and glycosidic conformations were identical to that of B-DNA. These results showed that W-C base pairing is not limited by the diameter of the helix. CD and  $T_m$  analysis also showed that xDNA, with its increased helical diameter, formed a thermodynamically stable helix.



**Figure 111..** Examples of x-modified modified nucleosides.

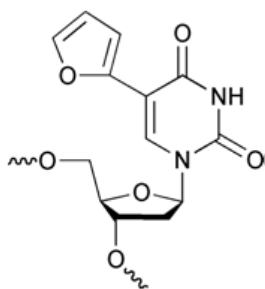
yDNA which differs from xDNA only in the direction of expansion, also showed remarkable helix stability compared to DNA controls. A naturally occurring octameric sequence of DNA displayed greater duplex stability with a complementary sequence composed entirely of yDNA compared to the natural complementary DNA strand. yyDNA (Figure 1.12) , the set of fluorescent naphtho-homologated DNA nucleobases, designed to test the true limits of W-C base pairing also formed stable helices. Observations such as these have led to increasingly accurate models of the structure of

DNA and the nature of Watson-Crick base pairing and have made unnatural nucleic acids an indispensable tool in the scientific field.



**Figure 1.12.** Examples of modified modified nucleosides.

Other fluorescent nucleobase analogs have proven to be an indispensable tool in biochemistry and molecular biology. Greco and Tor have synthesized a fluorescent pyrimidine nucleobase which when paired with a DNA abasic site, displays significant fluorescence (Figure 1.13)<sup>28</sup>.

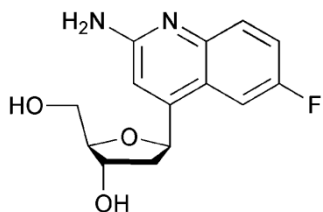


**Figure 1.13.** Fluorescent Pyrimidine Analog.

In addition those that recognize complementary single stranded RNA and DNA, nucleoside analogs that target double stranded DNA (dsDNA) have been the subject of much attention. Using phosphoramidites prepared from monomers including and similar to the molecule shown in Figure 1.14, Gold and co-workers have been able to synthesize



these triplex-forming oligonucleotides (TFOs) that specifically recognize T:A, A:T,G:C or C:G base-pairs via Hoogsteen base pairing<sup>29</sup>. These TFOs potentially allow for control of gene expression at the transcriptional level by binding to sequence-specific regions on the DNA duplex.



**Figure 1.14.** Gold's nucleoside analog designed target A:T base pair.

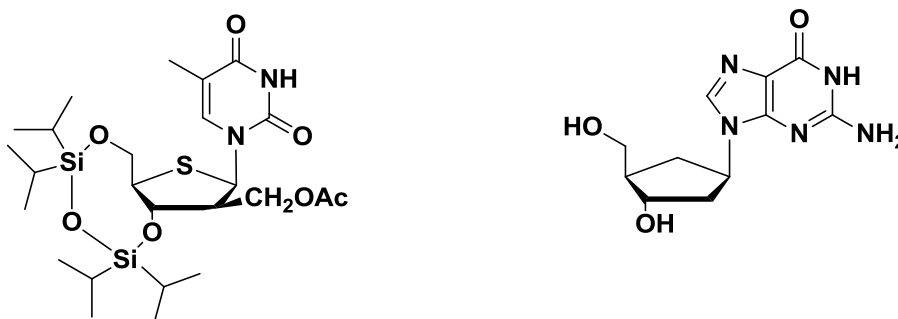
#### **1.4.2. Sugar Modification**

Manipulation of the ribose or deoxyribose portion of the nucleic acid also confers oligonucleotide stability towards nuclease degradation. Additionally, some of the truly ingenious modifications that are described below have allowed for improved plasmid DNA functionality<sup>30</sup>, DNA triplex formation<sup>31</sup> and nucleic acid – protein binding<sup>32</sup>.

A common modification is the substitution of the ring oxygen with sulfur. Haraguchi and co-workers recently devised the synthesis of the 4'-thionucleoside (Figure 1.15)<sup>33</sup>. When incorporated into an oligonucleotide, this derivative showed marked resistance to 3'-exonuclease cleavage. Interestingly, the 4'-thioDNA:4'-thioDNA homoduplex adopted A-form characteristics under physiological conditions. The affinity of this oligonucleotide for a complementary RNA sequence was not tested; however, the RNA-

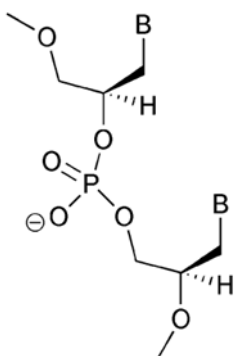
like features of the homoduplex suggest an increased affinity for the complementary RNA sequence compared to the native oligonucleotide.

A series of carbocyclic (deoxy)ribose nucleoside derivatives have also been synthesized. These nucleosides have had the ring oxygen replaced by a carbon atom. Carell and co-workers exploited the inability of these carbocyclic nucleosides to anomerize in order to study DNA lesions<sup>34</sup>. Previously, a rigorous study of such lesions was not possible since base-pairing effects could not be quantified due to the rapid inter-conversion of the  $\alpha$  and  $\beta$  conformations. Carbocyclic nucleosides have also been synthesized to act as agonists for A3 adenosine receptors which have critical cardiac and immunological functions<sup>35</sup>.



**Figure 1.15.** Left- Haraguchi's 4'-thionucleoside. Right- Carell's carbocyclic guanine.

Attempts to completely replace the ribose structure have also been made. In an experiment that test the limits of Watson-Crick base pairing, Meggers and co-workers have synthesized a nucleoside derivative that completely replaces the ribose sugar with a 3-carbon glycol framework (Figure 1.16)<sup>36</sup>. When incorporated into oligonucleotides, these simplified nucleic acids form stable duplexes - suggesting a possible predecessor to RNA as the genetic material for early life.

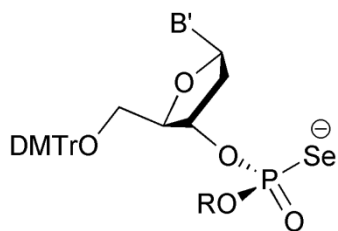


**Figure 1.16.** Megger's glycol-based oligonucleotide.

Another class of modified nucleosides is rigidified nucleosides. In these classes of molecules the nucleosides are rigidified by restricting the sugar or the base with respect to the sugar. We will discuss about it in Chapter 3.

#### **1.4.3. Backbone Modification**

While modification of the nucleobase has allowed for the exploration of the limits of W-C base pairing, the expansion of the genetic code and the ability to control gene expression at the transcriptional level, the susceptibility of unnatural oligonucleotides to enzymatic degradation is linked more closely to the nucleic acid backbone. A common modification of the phosphodiester bridge is the replacement of a non-bridging oxygen with a sulfur atom. Not only does this modification increase stability to endogenous enzymes, it enhances protein binding, reduces clearance and improves bioavailability. Stec and co-workers have synthesized a related phosphoroselenoate linker (Figure 1.17) and the Rozners lab has prepared oligoribonucleotides with amide linkages<sup>37</sup>.

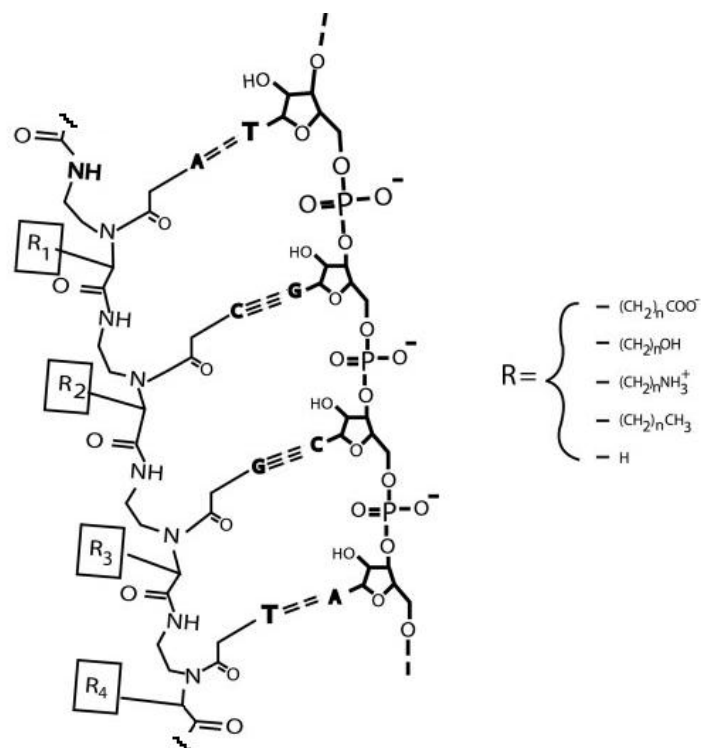


**Figure 1.17.** Stec's phosphoroselenoate linker.

Another important modification is to replace the negatively charged phosphodiester backbone by peptide linkage. The widely research polymers in this category are Peptide Nucleic Acids (PNA). We will discuss about it in the next section.

### 1.5. Peptide Nucleic Acid (PNA)

PNA (Figure 1.18) was originally designed as a ligand for the recognition of double stranded DNA<sup>38</sup>. The concept was to mimic an oligonucleotide binding to double stranded DNA via Hoogsteen base pairing in the major groove<sup>39</sup>. Thus the nucleobases of DNA were retained, but the deoxyribose phosphodiester backbone of DNA was replaced by a pseudo-peptide backbone. In theory a neutral (peptide) backbone should improve the triplex binding capability of the ligand, and it is believed that the pseudo-peptide backbone was a good chemical scaffold that would allow us to design recognition moieties that went beyond homopurine targets. There has been a wide variety of substitutions in the backbone, that we will discuss about later.



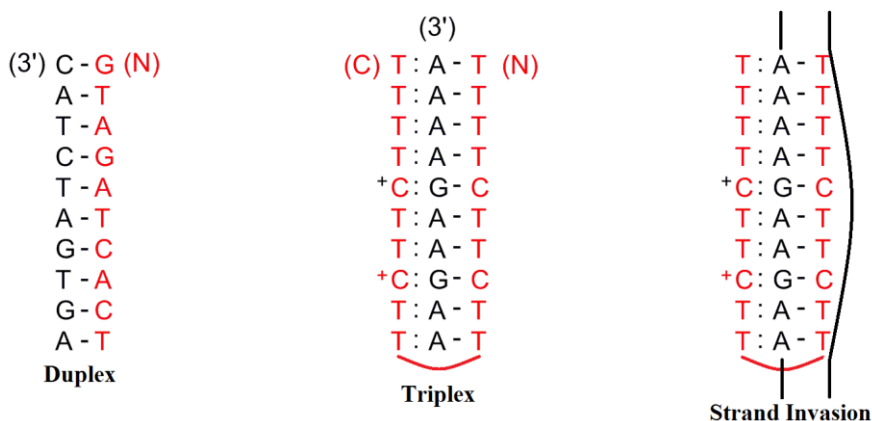
**Figure 1.18.** Illustration of a PNA-RNA duplex with possible substitutions in the backbone of PNA.

PNA as we know them today have a backbone made from repeating N-(2-aminoethyl)glycine units linked by peptide bonds<sup>40</sup>. The different bases (purines and pyrimidines) are joined to the backbone by methylene carbonyl linkages. They are depicted as peptides from the N terminus to the C terminus, corresponding to the 5' to 3' direction as in DNA because all intramolecular distances and the configuration of the nucleobases are similar to those occurs between PNA and DNA or RNA sequences by hydrogen bonding (Figure 1.18). The uncharged nature of PNA is responsible for the better thermal stability of PNA/DNA duplexes compared with their DNA/DNA equivalents.

### 1.5.1. Properties of PNA

#### Modes of Binding

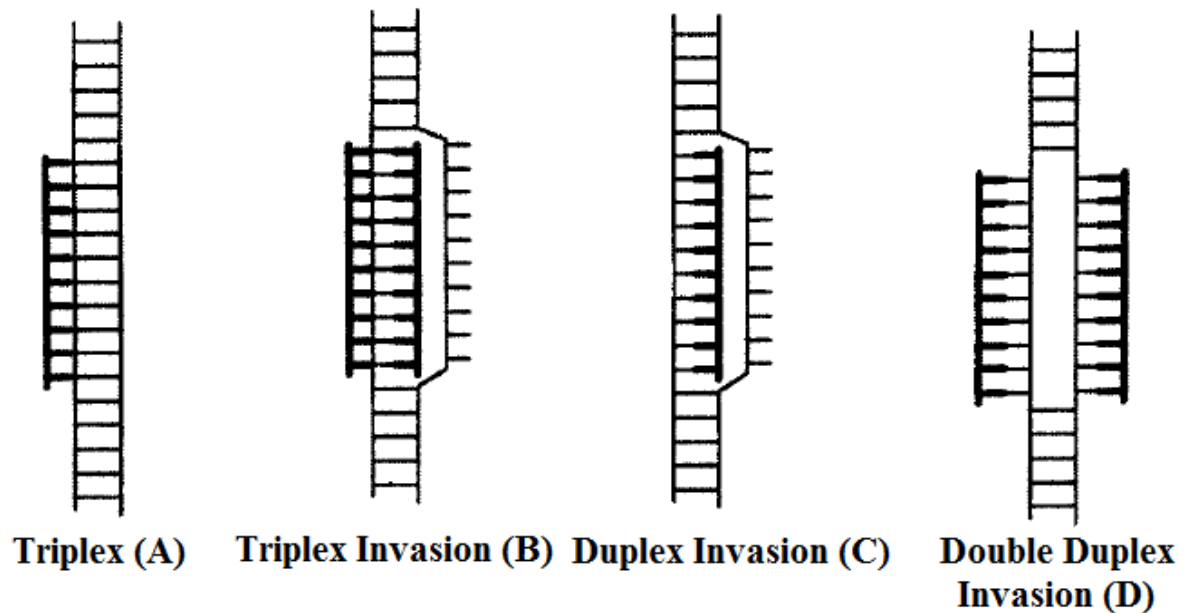
It is important to emphasize the directionality of PNA while binding to DNA (Figure 1.19).



**Figure 1.19.** Schematic representation of a PNA-DNA duplex in the antiparallel mode (3'-end of the DNA facing the amino-terminal of the PNA), a 2PNA/DNA triplex in the preferred binding mode with antiparallel Watson-Crick strand and parallel Hoogsteen strand a linker of a bis-PNA is indicated), and a triplex strand invasion complex using a bis-PNA. PNAs are shown in red. Watson-Crick hydrogen bonds are indicated by '-', while Hoogsteen hydrogen bonds are indicated by ':'.

There are four different binding modes on targeting double-stranded DNA with PNAs have been described (Figure 1.20). These include conventional triplex binding in the major groove of the DNA double helix<sup>41</sup> (Figure 1.20 A) and triplex invasion in which two PNA oligomers invade and open this DNA double helix through formation of an internal PNA-DNA-PNA Watson-Crick-Hoogsteen triplex<sup>40</sup> (Figure 1.20 B and C). Duplex invasion through an internal PNA-DNA Watson-Crick duplex<sup>42</sup> (Figure 1.20 C) has also been observed under certain conditions, and if special modified pseudo-complementary bases, such as the diaminopurine-thiouracil base pair, are employed in

the PNA oligomers, binding to the target by double duplex invasion is possible (Figure 1.20 D).



**Figure 1.20.** Illustration of the binding modes on targeting DNA duplex by PNA. The ladder represents the DNA duplex and the PNA oligomer is bolded.

### ***Structure of PNA Complexes***

Three-dimensional molecular structures have been determined for the basic PNA complexes. The structures of a 9-mer PNA<sub>2</sub>-DNA triplex<sup>43</sup> and a hexamer PNA-PNA duplex<sup>44</sup> were determined by X-ray crystallography; while the structures of a 6-mer PNA-RNA<sup>45</sup> and an 8-mer PNA-DNA<sup>46</sup> duplex were determined by NMR. Following are the observations about the structure of PNA hybrids:

- The binding of PNA to single stranded DNA or RNA allows the backbone of the nucleic acid partner to retain close to its ‘natural’ conformation. The RNA in the

PNA-RNA duplex is close to an A-form with C3'-*endo* sugar puckering; whereas the DNA in the PNA-DNA duplex is closer to B-form with C2'-*endo* sugar puckering.

- The two duplexes also show a tendency towards a larger pitch than the usual 10-11 base pairs seen in nucleic acid duplexes.
- Both the PNA<sub>2</sub>-DNA triplex and especially the PNA-PNA duplex both show a very wide (26 and 28 Å, respectively) and a very large pitched helix (16 and 18 bases for the triplex and duplex respectively)

### ***Thermal Stability of PNA hybrid complexes***

One of the most important aspects of the PNA hybrid complexes is the higher thermal stability of PNA/DNA and PNA/RNA duplexes compared with DNA/DNA and DNA/RNA duplexes<sup>47</sup>. This stronger binding is attributed to the lack of charge repulsion between the neutral PNA strand and the DNA or RNA strand. As a general rule,  $T_m$  of a PNA/DNA duplex is 1 °C higher per base pair than  $T_m$  of the corresponding DNA/DNA duplex (in 100 mmol/dm<sup>3</sup> NaCl). Some of the following data show the greater stability (reflected in the higher  $T_m$  values) of PNA/DNA or PNA/RNA duplexes<sup>48</sup>: a) 15-mer PNA/DNA,  $T_m$  = 69 °C; b) 15-mer DNA/DNA,  $T_m$  = 54 °C; c) 15-mer PNA/RNA,  $T_m$  = 72 °C; d) 15-mer DNA/RNA,  $T_m$  = 50 °C.

### ***Effect of Salt Concentration on Binding***

The  $T_m$  values of PNA/DNA duplexes are practically independent of salt concentration. This is striking difference to that of the DNA/DNA duplexes, which are highly dependent on ionic strength. At low ionic strength, PNA can bind to a target sequence at



temperatures where DNA hybridization is disfavored. Hybridization of PNA is also fairly independent of the concentration of magnesium ion<sup>40</sup>.

### ***Binding Specificity***

PNA also shows greater specificity in binding to complementary DNA<sup>49</sup>. A PNA/DNA mismatch is more destabilizing than a mismatch in a DNA/DNA duplex. A single mismatch in mixed PNA/DNA 15-mers lowers  $T_m$  by 8–20 °C (15°C on average). In the corresponding DNA/DNA duplexes a single mismatch lowers  $T_m$  by 4–16 °C (11°C on average).

### ***Resistance to nucleases and proteases***

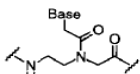
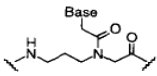
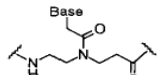
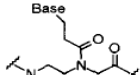
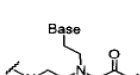
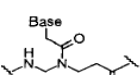
PNAs with their peptide backbone bearing purine and pyrimidine bases are not molecular species that are easily recognizable by either nucleases or proteases<sup>50</sup>. Therefore the lifetime of these compounds is extended both *in vivo* and *in vitro*.

### ***PNA Insolubility in Water***

Being charge-neutral biopolymers, PNA have poor water solubility compared with DNA or RNA. PNA molecules have a tendency to aggregate to a degree that is dependent on the ratio of purine to pyrimidine, the sequence and length of the oligomer<sup>42,47</sup>. Some recent modifications, including the incorporation of positively charged lysine residues (carboxyl-terminal or backbone modification in place of glycine), have shown improvements in solubility.

### 1.5.2. Modification of PNA Backbone

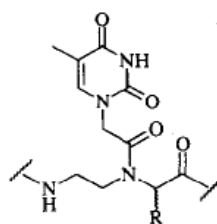
There have been two major directions to modify the backbone to enhance its binding properties and solubility in water. First, change the length of the backbone and the length between the backbone and the base. Second, substitutions at the  $\alpha$ -position of the glycine moiety. Extensions in any of the three possible directions<sup>51</sup> (entry 2-4, Table 1.2) are not desirable, and the restricted flexibility imposed by the secondary amido-group also seems necessary since reduction of this to a tertiary amine (entry 5, Table 1.2) is deleterious to the hybridization potency<sup>52</sup>. Also, an isomerization of the backbone to the ‘retro-inverso’ structure (entry 6, Table 1.2) which essentially moves a methylene group from the ‘ethyl’ to the ‘glycine’ moiety results in PNAs of low hybridization efficiency.

Entry	Structure	Backbone/linker	$\Delta T_m$ DNA	$\Delta T_m$ RNA
1		ethylglycine	0	0
2		propylglycine	-8.0	-6.5
3		ethyl- $\beta$ -alanine	-10	-7.5
4		propionyl linker	-20	-16
5		ethyl linker	-22	-18
6		retro-inverso	-6.5	nd

**Table 1.2.** Effects on thermal stability ( $\Delta T_m$  °C) for structurally modified PNA in which T-monomers when incorporated into the oligomer sequence H-GTA GAT CAC T-NH<sub>2</sub><sup>53</sup>.

On the other hand, substitutions at  $\alpha$ -position of the glycine unit in the backbone show some promise. Table 1.3 summarizes some the results obtained by this strategy. It is

evident that PNA backbones based on other natural amino acids than glycine result in fair to good DNA mimicking properties with some variations amongst the different amino acids<sup>54</sup>. By using such PNA monomers, various chemical scaffolds can be introduced into the backbone and the physical properties of the PNAs, such as solubility, hydrophobicity, ionic properties etc. can be modulated. Especially it was shown that PNAs containing a few lysine-monomers (entry 14 and 15, Table 1.3) have increased aqueous solubility.



Entry	R	chirality	$\Delta T_m$ DNA	$\Delta T_m$ RNA
7	CH <sub>3</sub>	L	-1.8	nd
8	CH <sub>3</sub>	D	-0.7	nd
9	<i>sec</i> -Bu	L	-2.6	-3.0
10	CH <sub>2</sub> OH	L	-1.0	-1.0
11	CH <sub>2</sub> OH	D	-0.6	-1.0
12	CH <sub>2</sub> CO <sub>2</sub> H	L	-3.3	nd
13	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> H	D	-2.3	nd
14	(CH <sub>2</sub> ) <sub>4</sub> NH <sub>2</sub>	L	-1.0	-1.3
15	(CH <sub>2</sub> ) <sub>4</sub> NH <sub>2</sub>	D	+1.0	0

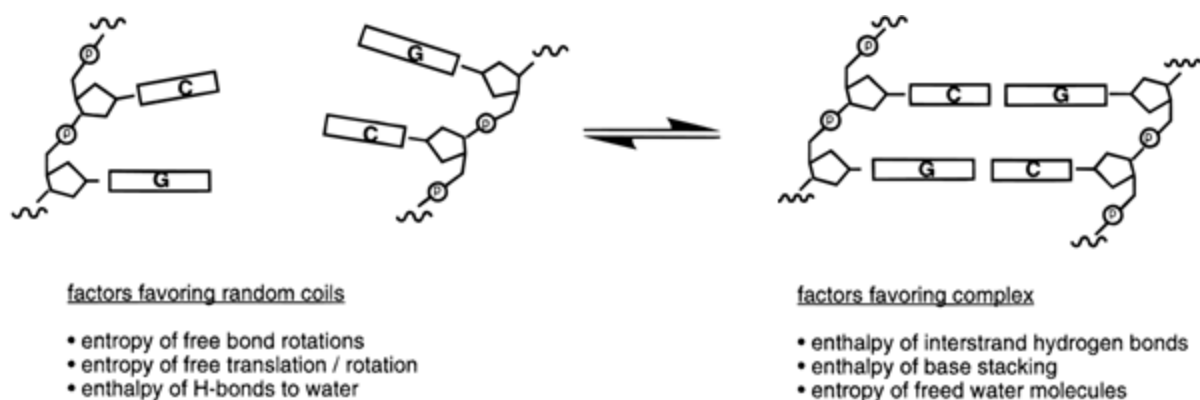
**Table 1.3.** Effects on thermal stability ( $\Delta T_m$  °C) for the PNA sequence H-GTA GAT CAC T-NH<sub>2</sub> incorporating three chiral monomers as compared to an unmodified PNA<sup>53</sup>.

## 1.6. Factors Determining the Modified Nucleic Acids Binding Properties

One of the primary goals to develop the modified nucleic acids is to form stronger binding to the complementary DNA or RNA. In principle it can be achieved either by increasing the enthalpy and/or decreasing the entropy of duplexation. Enthalpic stabilization of duplexation can be addressed by substituting the negatively charged phosphate groups in natural oligonucleotides with charge-neutral nucleoside linking groups such as phosphotriester<sup>55</sup>, methyl phosphonate<sup>56</sup>, phosphoramidate<sup>57</sup>, sulfonate<sup>58</sup>, and acetal groups<sup>59</sup>. Substitution of 2,6-diaminopurine for adenine as the nucleobase in DNA oligomers<sup>60</sup> stands as an example of enhanced enthalpic stabilization of the resulting duplexes due to an additional H-bond in the diaminopurine-thymine base pair with respect to the parent adenine-thymine base pair. We discussed some other examples in section 1.5. Entropic stabilization can be achieved by proper preorganization. Nucleosides have been designed and synthesized to confer, as constituents of an oligonucleotide chain, a higher degree of preorganization of a single strand for duplex formation with respect to natural DNA, thus leading to an entropic advantage for the pairing process.

The concept of “preorganization” was first introduced by Cram in the analysis and design of small organic guest-host complexes<sup>61</sup>. It was found that entropy factor disfavors the formation of such guest-host complexes for two reasons: i) three degrees of translational and rotational entropy are lost in forming a complex from two separate molecules, ii) if either of the separate molecules has free internal bond rotations, then the “freezing out” of all such rotations on forming the complex is also entropically unfavorable<sup>62</sup>. For molecular recognition, the first can’t be avoided because the phenomenon fundamentally

needs this to happen. However, the second factor can be addressed by the tuning, designing and testing. If a host molecule or ligand is constructed so as to be rigidly held in the binding conformation prior to complexation, then little or no entropic cost due to fixing bond rotations will be necessary in binding<sup>44</sup>. In other words, a flexible molecule can be preorganized in the synthesis for optimum binding. It is to be emphasized here that the preorganization should not hold the molecule rigidly in the wrong conformation<sup>63</sup>.

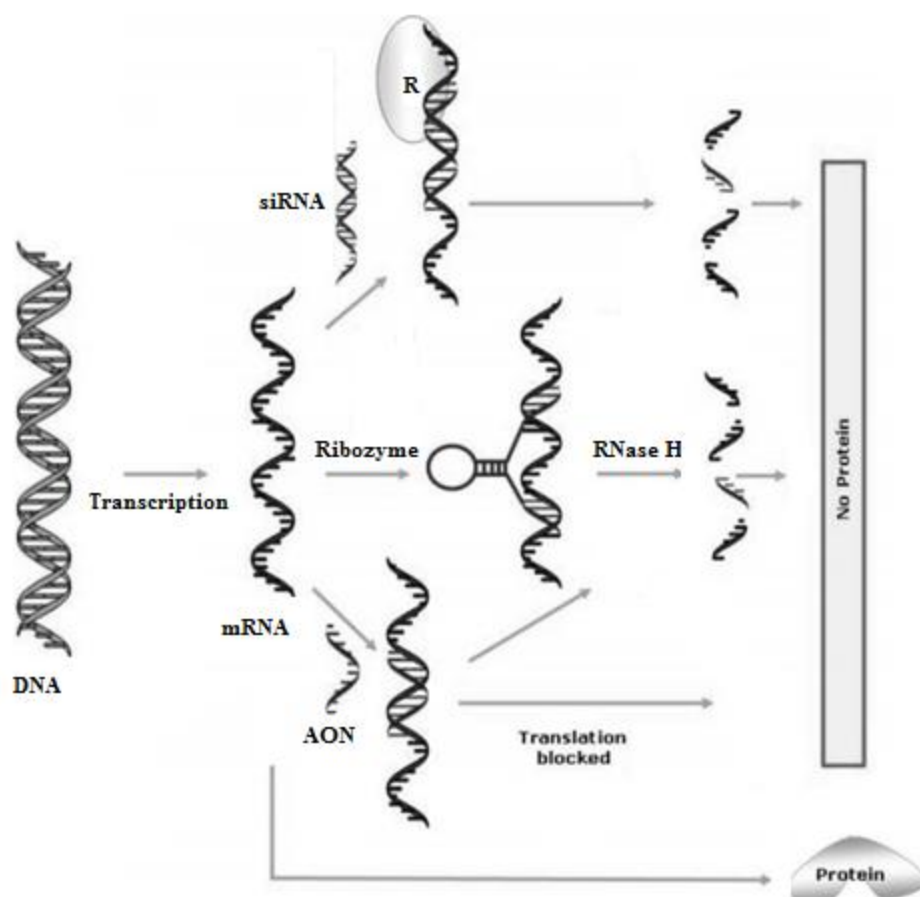


**Figure 1.21.** Competitive thermodynamic factors favoring the complex formation.

The concept of preorganization holds true for nucleic acid duplex formation, especially sequence specific recognition of a single-stranded nucleic acid by another oligonucleotide. A single-stranded nucleic acid is quite flexible, with rotation of the nucleobase with respect to the sugar moiety about the glycosidic bond and about five freely rotating bonds per phosphate. Helix formation is therefore highly unfavorable entropically and is a spontaneous reaction only because a very favorable enthalpy term just compensates (Figure 1.21).

### 1.7. Potential Applications of Modified Oligonucleotides as Therapeutics

Modified nucleotides have found applications in therapeutics by modulating the gene expression. Broadly, an oligonucleotide designed to alter gene expression can act either at the level of DNA duplex to inhibit transcription (also known as “the antigene approach”) or at the level of m-RNA to inhibit translation (known as “the antisense approach”). The overall mechanism has been summarized in Figure 1.22.



**Figure 1.22.** Conceptual representation of antisense-mediated gene-silencing approaches (R represents RNA-induced silencing complex, or RISC).

### 1.7.1. Perturbing Gene Expression at Transcriptional Level (Antigene Approach)

By hybridizing synthetic oligodeoxynucleotides with double-stranded DNA<sup>64</sup>- These hybrids are generally formed within the major groove of the helix, but there are a few strategies for hybridization within the minor groove that have also been reported<sup>65</sup>. In either case, a triple-stranded molecule is produced, hence the origin of the term triple-helix forming oligodeoxynucleotide (TFO) (Figure 1.22 and 1.23). TFOs may act in two ways. They may prevent binding of transcription factors to the gene's promoter and therefore inhibit transcription. Alternatively, they may prevent duplex unwinding and, therefore, transcription of genes within the triple-helical structure.



**Figure 1.23.** Triplex Forming Oligonucleotide (TFO) in the Major Groove of DNA Duplex.

*Use of specific nucleic acid sequences to act as “decoys” for transcription factors<sup>66</sup>-* Since transcription factor proteins recognize and bind specific DNA sequences, it is possible to synthesize nucleic acids that will effectively compete with the native DNA sequences for available transcription factor proteins *in vivo*. However, in practice this approach has limitations. For example, unless single gene transcription factors can be identified, it is difficult to conceive how this approach, though potentially effective for controlling cell growth, can be made gene specific.

### ***1.7.2. Perturbing Gene Expression at Translational Level***

*Antisense Approach-* In antisense strategy of gene silencing, the oligonucleotide forms a stable duplex with the targeted m-RNA and inhibits translation either by blocking the binding and assembly of the translational machinery to the m-RNA or cleaves the target mRNA by recruitment of RNase H<sup>67</sup>. Steric blockade of the translation is achieved when an antisense oligonucleotide is directed to the 5'-terminus of the target, while RNase H-mediated cleavage occurs when the antisense oligonucleotide binds to the target strand at any site to produce a duplex conformation of a DNA-RNA heteroduplex. This heteroduplex motif serves as a recognition element of RNase H recruitment, leading to the cleavage of the RNA strand<sup>68</sup>. Inhibition of translation can also be achieved by small interfering RNA (siRNA). siRNAs result when exogenous agents like viruses and transposons or endogenous genes introduce long double-stranded RNA (dsRNA triggers gene silencing by a process which is widely used nowadays known as RNA interference (RNAi) (Figure 1.22).

### **1.8. Application of Modified Oligonucleotides as Diagnostic Tool**

*Probes for Single Nucleotide Polymorphism (SNP) Detection-* SNPs are the DNA sequence variations occurring when a single nucleotide in the genome differs between members of a species. SNPs may fall within coding sequences of genes, noncoding regions that may have regulatory effects, or the regions between the genes. SNPs in the coding region may alter the gene product (polypeptide) and cause several ailments. The high binding affinity and sequence specificity of modified nucleotides find significant use



in diagnostics where they are used as probes in several hybridization-based assays such as expression profiling, DNA sequencing, SNP genotyping, and so forth<sup>69</sup>.

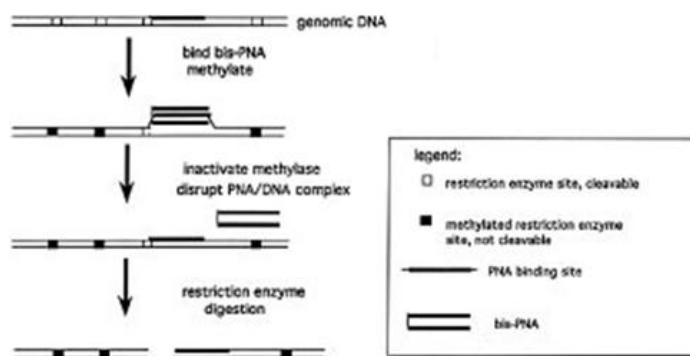
***Probes for RNA Capture-*** For studying expression level and localization of RNA within all types of eukaryotic cells, *in situ* detection of RNAs by hybridization with complementary probes is a powerful technique. The favorable properties of the modified oligonucleotides make them a good candidate to be used as probes for RNA detection<sup>70</sup>.

***Probes for micro-RNA detection (miRNA)-*** miRNA are small (20-23 nucleotide) noncoding RNAs that regulate gene expression by base-pairing to partially complementary mRNAs. The detection and characterization of miRNAs have been technically challenging because of their small size. Modified oligonucleotides, especially DNA chimera with several positions substituted by LNA residues have been used to significantly increase detection of low-abundance miRNAs<sup>71</sup>.

## **1.9. Applications of PNA**

Apart from its potential use as therapeutics to selective inhibition at the transcriptional and translational level (Section 1.7) because of their exceptional binding properties, they can be used as biomolecular tools. Some of their applications, not an exhaustive list, are mentioned below:

## *PNA Assisted Rare Cleavage*<sup>72</sup>



**Figure 1.24.** Schematic representation of major steps in PNA Assisted Rare Cleavage.

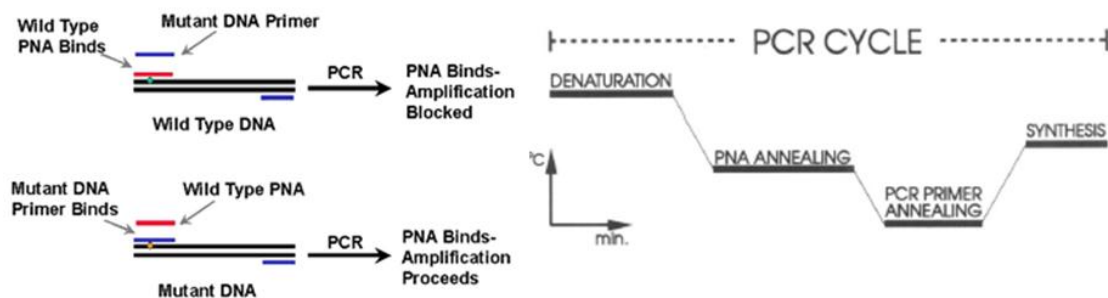
It is based on the general “Achilles’ heel” cleavage strategy. This methodology makes it possible to convert usual restriction enzyme into infrequent genome cutters. In this method (Figure 1.24), a very stable and sequence-specific complex is formed between double-stranded genomic DNA and a cationic pyrimidine bis-PNA. Then the sample is treated with methylases, the bis-PNA is removed from the DNA and the sample is treated with a restriction enzyme. The enzyme recognizes the same sites as methylase and thus cannot cleave them. The only exceptions are the very few non-methylated sites, which are protected against methylation by the bis-PNA. These rare sites become accessible for enzymatic recognition after PNA is removed. As a result, the restriction enzyme cuts the genomic DNA into small number of fragments. A pool of numerous combinations of various bis-PNAs with different methylation/restriction enzymatic pairs generates a new class of genome rare cutters.

### ***PNA for Artificial Restriction System***

There are many sequences which are not recognized by the restriction enzymes; in such cases, PNAs are designed to have their oligonucleotides complementary to the recognition sequences. These PNAs hybridize to the complementary targets on dsDNA and loop out the non-complementary DNA sequences<sup>73</sup>. The single-stranded looped-out fragments of DNA are then removed by S1 nuclease, which cleaves single-stranded nucleic acids. Therefore PNA can be used as a cutting tool in combination with S1 nuclease to make an ‘artificial restriction enzyme’ system.

### ***PCR Clamping***<sup>74</sup>

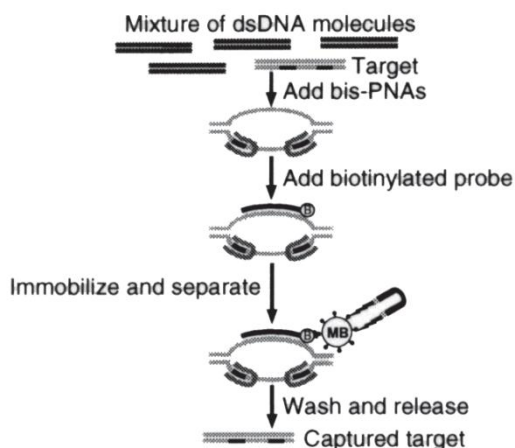
This method enhances the specificity of the PCR reaction by targeting the initial step involved in non-specific amplification<sup>75</sup>, i.e., the binding of the primer to a mis-matched target sequence. As shown in Figure 1, the method operates by competition for a common target site between a PNA (complementary to the wild-type target sequence) and one of the PCR primers (complementary to the mutant target sequence), or vice versa. When the template contains the wild-type sequence, PNA binding will dominate over primer binding due to the higher affinity of the matched PNA for the target site. As PNA cannot be extended by the Taq-polymerase the effect of this binding is that the amplification reaction is impaired. When the mutant sequence is present, PCR primer binding will dominate over PNA binding with the resulting generation of amplicons (Figure 1.25).



**Figure 1.25.** Schematic illustration of PCR Clamping. Left- Steps; Right-Typical PCR Cycle.

### ***Oligonucleotide/PNA-assisted Affinity Capture (OPAC)***<sup>76</sup>

It is a method of sequence-specific isolation and purification of intact double-stranded DNA (dsDNA). The protocol involves three steps. First, two cationic bis-PNAs locally pry the DNA duplex apart at a predetermined site. Then, the exposed DNA single strand is targeted by a complementary biotinylated ODN to selectively form a stable PD-loop complex. Finally, the capture of dsDNA is performed using streptavidin covered magnetic beads. In essence, the methodology is based on selective tagging of a DNA duplex by biotinylated oligodeoxynucleotide through formation of a so-called PD-loop by the aid of a pair of PNA “openers (Figure 1.26).



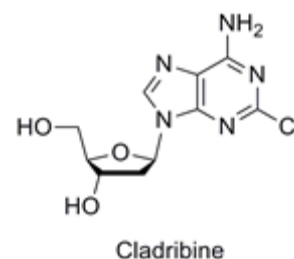
**Figure 1.26.** Schematic illustration of steps involved in OPAC methodology.

### 1.10. Specific Examples of Modified Nucleosides Which Have Medicinal Applications

Some modified nucleosides are already approved by the FDA for medicinal applications whereas others are under investigation for their potentiality as drugs. A discussion of a few examples can be found below.

#### *Cladribine*

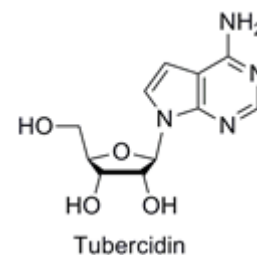
As a purine analogue, it is a synthetic anti-cancer agent that also suppresses the immune system. Chemically, it mimics the nucleoside adenosine and inhibits the enzyme adenosine deaminase, which interferes with the cell's ability to process DNA. It is easily destroyed by



normal cells except for blood cells, with the result that it produces relatively few side effects and results in very little non-target cell loss. It is approved for the treatment of symptomatic hairy cell leukemia<sup>77</sup>.

#### *Tubercidin*

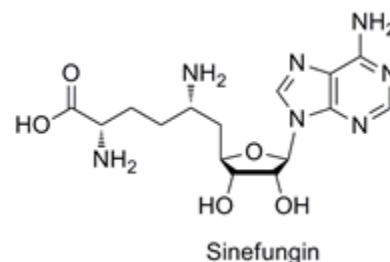
It is a toxic adenosine analogue with antiviral, anti-trypanosomal, and antifungal functions. It inhibits multiple metabolic processes which include: RNA processing, nucleic acid synthesis, protein synthesis, and methylation of tRNA through intracellular incorporation into nucleic



acids. It also acts as a plant antifungal, inhibits mammalian SAH hydrolase (SAHH), and blocks purine biosynthesis in *Candida famata*. And many others have been developed by chemists with specific goals in mind<sup>78</sup>.

### ***Sinefungin***

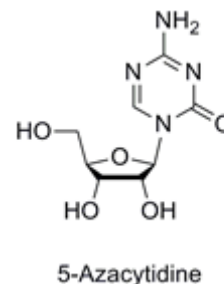
Sinefungin is a DNA methyltransferase inhibitor and an analog of *S*-adenosyl-L-methionine (SAM) as indicated by previous studies. In studies investigating cytosine deamination by DNA methyltransferase, it was noted that the presence of Sinefungin



increased the rate of deamination 103 fold in the presence of its analog SAM, and 104 in the absence of SAM. This may suggest that Sinefungin could be useful as a test compound in antiviral, anticancer, antiparasitic and antifungal applications. Additional studies indicate that Sinefungin was able to inhibit RNA synthesis in a vaccinia virion preparation<sup>79</sup>.

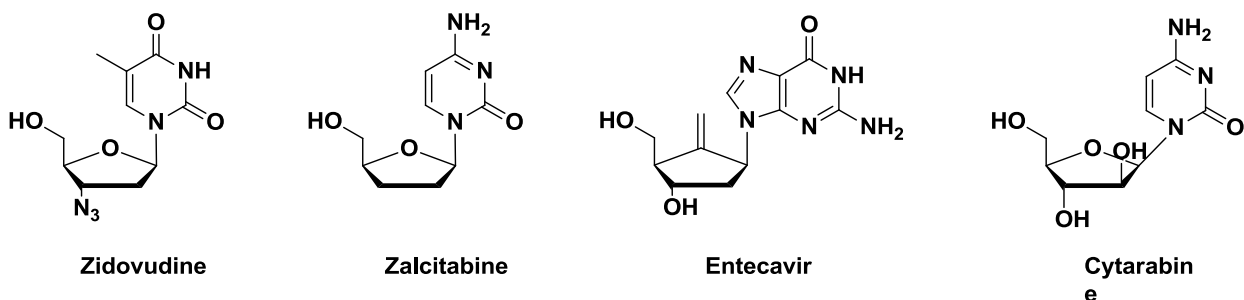
### ***5-Azacytidine***

It is a potent growth inhibitor and cytotoxic agent and a potent DNA methyltransferase inhibitor. It causes DNA demethylation or hemi-demethylation, creating openings that allow transcription factors to bind to DNA and reactivate tumor suppressor genes. It has also recently been shown to increase reprogramming efficiency of stem cells 10-fold<sup>80</sup>.



### ***Significant other examples***

Chain terminators, such as Zidovudine or Zalcitabine, lack the 3'-hydroxyl necessary for addition of sequential nucleosides during replication. Other modifications include those seen in Entecavir and Cytarabine. Entecavir has been shown to block the active site of viral DNA polymerase.



Cytarabine is used as a chemotherapeutic, but works in the same principal by blocking DNA and RNA polymerases in rapidly dividing cells.

### 1.11. Scope of the Research

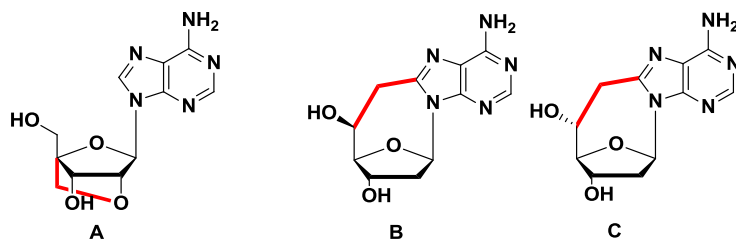
McLaughlin's group focuses primarily on the chemical modification of nucleosides followed by their incorporation in the nucleic acid strand to study their biophysical and biochemical properties. Enzymatic incorporation of the triphosphates of these modified nucleosides provides another direction for investigation. As we have discussed, structurally a nucleic acid strand has three different moieties- the nucleobase, the sugar and the phosphodiester backbone (Figure 1.1). Our group develops novel nucleosides by modifying some combination of the three. One of the projects focuses on restricting the free rotation of the nucleobase about the glycosidic bond to the anti-conformation. This is achieved by tethering the C8 of the purine or the C6 of the pyrimidine with the C5' of the sugar. Another major project is to perform minor groove modifications. These are performed by introducing different substituents at the C3 position of 3-deaza-purines. These modifications provide tools to study the minor groove, including, but not limited to, the study of the role the spine of hydration plays in the stability of DNA duplexes.

One of the strategies to prepare these modified nucleosides is to employ glycosylation methodology to couple the sugar with the nucleobase component. Previous reported procedures did not afford the desired results, in terms of regio- and stereo-selectivity of the glycosylation<sup>81</sup>. This inspired us to develop methodology to perform this step more efficiently. It is not affordable to carry out glycosylations that yield only 25-40% of the desired  $\beta$ -N9 product, or sometime even lower overall yield, considering the fact that it takes 8-10 difficult steps to synthesize these modified bases. Glycosylation reactions via sodium salt method have a number of issues. In my first project, I have tried to address some of these issues. Previous research in our lab used tetramethylsuccinic anhydride as a directing/protecting group in purine glycosylations<sup>82</sup>. This methodology proved that a cyclic diimide with sufficient steric bulk can function as a directing/protecting group. With that observation confirmed, we sought to address whether increasing the steric bulk at the sites of the methyl groups would further enhance the directing ability of substituted succinimide derivatives. Further investigation resulted in the development of bicyclo-1,1'-dicarboxylic acid cyclic-anhydride as a directing/protecting, which established our hypothesis about sterics at C6 position of the purines.

Glycosylation reactions are not the only method to access the modified nucleosides. In my next project, I developed 8, 6'-(*S*)- and 8, 6'-(*R*)-cyclo-2'deoxyadenosine (Figure 1.28 B and C). The inspiration for developing the route to access these epimers came from the amazing binding properties of Locked Nucleic Acids (LNA)<sup>83</sup> (Figure 1.27 A) to native nucleic acids, especially RNAs. The monomers of the LNA is locked in the C3' - *endo* sugar conformation due to rigidity in the structure introduced by the tethering C2' - oxygen and C4' of the sugar moiety via methylene group. It has been reasoned that due to



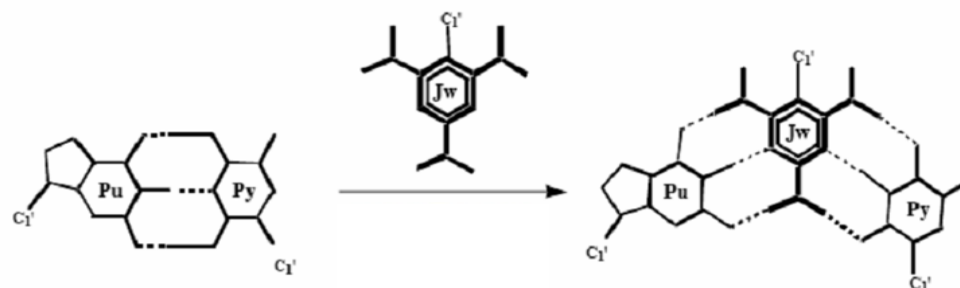
the rigidity of the monomers, when they are introduced in the nucleic acid strands they are pre-organized to form A-type helices. This preorganization is responsible for the high binding selectivity and specificity properties. With that established approach, we asked whether we could introduce rigidity in the structure of the nucleoside monomers by freezing the free rotation of the nucleobase about the sugar moiety. This has been achieved by introducing intra-residue covalent bond between C5' and C8 of purine or C5' and C6 of pyrimidine in the first generation analogues. In the second generation, the same two centers are connected via a methylene group. In my second project, we were able to optimize an efficient route to synthesize both the epimers at C5' position of 8, 5'-methylene-cyclo-2'-deoxyadenosine (Figure 1.27 B and C). The goal of my second project was to develop these modified nucleosides, so that they could be incorporated in a DNA strand to study their properties like binding propensity towards other nucleic acid strands, especially RNAs etc.



**Figure 1.27.** (A) Monomer of LNA; (B) and (C) Epimers of 8, 5'-methylene-cyclo-2'-deoxyadenosine.

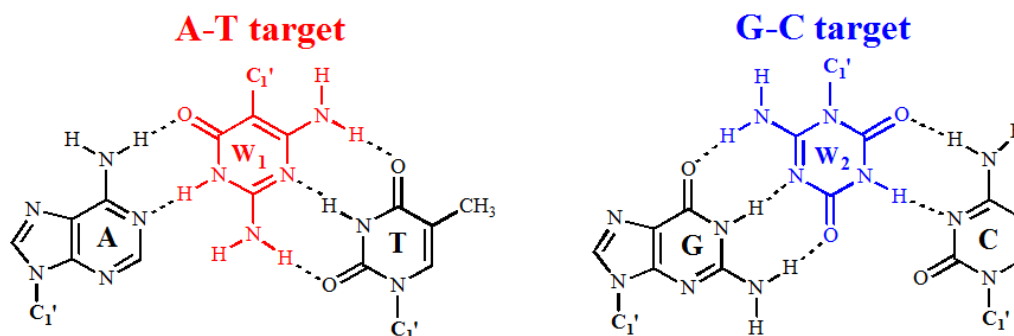
In my third project, the goal was to investigate whether we could sequence specifically recognize DNA duplexes by Janus Wedge format. The Janus-Wedge (Jw) triple helix<sup>84</sup> is a fundamentally new approach to target a DNA duplex. In this approach, an incoming

peptide nucleic acid (PNA) strand inserts itself between the Watson-Crick (W-C) faces of the base pairs in a duplex to generate a new (Jw) three-stranded complex (Figure 1.29)<sup>85</sup>.



**Figure 1.28.** Concept of Janus Wedge Triplex.

The bases on the PNA can be designed so that they have two hydrogen bond forming faces with correct hydrogen bond forming profile. The first generation J-W complexes (Figure 1.29) were based on a single ring format of a pyrimidinone ring (W1) for targeting A-T/T-A base pairs<sup>86</sup> and a triazinedione (W2) ring for targeting G-C/C-G base pairs<sup>87</sup>. However, specificity was a problem with the first generation bases. In my third project, we designed and synthesized second generation Jw residues to achieve the goal of specific DNA duplex targeting. These are asymmetric purine-like Jw base monomers that are designed to selectively target and discriminate all four W-C base pairs (A-T, T-A, C-G and G-C) and form the basis for the development of highly selective duplex DNA targeting modules.



**Figure 1.29.** First Generation Janus Wedge Bases.

---

## References

- <sup>1</sup> Saenger, W. *Principles of Nucleic Acid Structure*. Springer-Verlag Inc., New York **1984**.
- <sup>2</sup> a) Cohen, J. S. *Oligodeoxynucleotides: Antisense Inhibitors of Gene Expression*; CRC Press: Boca Raton, **1989**. b) Baserga, R., Denhardt, D. T. *Antisense Strategies*. *Ann. N.Y. Acad. Sci.***1992**, 660, 1. c) Milligan, J. F.; Matteucci, M. D.; Martin, J. C. *J. Med. Chem.***1993**, 36, 1923.
- <sup>3</sup> Dahm, R *Hum. Genet.***2008**, 122 (6), 565–581.
- <sup>4</sup> Jones, M. E. *Yale Journal of Biology and Medicine* **1953** 26 (1), 80–97.
- <sup>5</sup> Levene, P. *J Biol. Chem.* **1919**, 40 (2), 415–24.
- <sup>6</sup> Valery, N. S. *Nature Reviews Genetics* **2001**, 2 (9), 723–729
- <sup>7</sup> Lorenz, M. G.; Wackernagel, W. *Microbiol. Rev.* **1994**, 58 (3), 563–602.
- <sup>8</sup> Hershey, A.; Chase, M. *J Gen Physiol* **1952**, 36 (1), 39–56.
- <sup>9</sup> Chargaff E, Lipshitz R, Green C, Hodes ME. *J Biol Chem.* **1951**, 192 (1), 223–230.
- <sup>10</sup> Astbury, W. *Symp. SOC. Exp. Biol.* **1947**, 1 (66).
- <sup>11</sup> Franklin, R. E.; Gosling, R. G. *Nature*, **1953**, 172 (4369), 156–157.

- 
- <sup>12</sup> Westheimer, F. *Science* **1987**, *235*, 1173-1178.
- <sup>13</sup> a) Pinnow, J. Z. *Electrochem.* 1918, *24*, 21. b) Knowics, F. C.; Benson, A. W. *Trends Biochem. Sci.* **1983**, *8*, 178. c) Crafts, J. –M. *Bull. Soc. Chim. France Ser.2* **1870**, *14*, 99. d) Baer, C. D.; Edwards, J. O.; Rieger, P. H. *Inorg. Chem.* **1981**, *20*, 905. e) Aelion, R.; Loebel, A.; Eirich, F. *Recl. Trav. Chim. Pays-Bas* **1950**, *69*, 61; *J. Am. Chem. Soc.* **1950**, *72*, 5705.
- <sup>14</sup> Adopted from- Lu, X. J. and Olson, W. K. *Nucl. Acids Res.* **2003**, *31*(17), 5108-5121.
- <sup>15</sup> Pabo, C.; Sauer, R. *Annu Rev Biochem* **1984**, *53*, 293–321.
- <sup>16</sup> Ha, S. C., Lowenhaupt, K.; Rich, A., Kim, Y. G., Kim, K. K. *Nature* **2005**, *437*(7062), 1183–1186. b) Rich, A., Zhang, S. *Nature Review Genetics* **2003**, *4*(7), 566–572.
- <sup>17</sup> Nelson, D. L.; Cox, M. M. in *Lehninger Principles of Biochemistry* Third Edition. Worth: New York, **2000**, 338.
- <sup>18</sup> a) Rich, A.; Norheim, A.; Wang, A. H. J. *Annual Review of Biochemistry* **1984**, *53* (1), 791-846. b) Ho, P. S. *Proc. Natl. Acad. Sci.* **1994**, *91* (20), 9549-9553.
- <sup>19</sup> Ts'o, P. O. P. *Basic Principles in Nucleic Acid Chemistry Academic Press, New York* **1974**, *Vol. 1*, 237-318.
- <sup>20</sup> Tan, Z. J. and Chen, S. J. *Biophys J.* **2006**, *90*(4), 1175–1190.
- <sup>21</sup> Hud, N. V., Sklenar, V., Feigon, J. *J. Mol. Bio.*, **1999**, *286* (3), 651.
- <sup>22</sup> Woods, K. K.; Lan, T.; McLaughlin, L. W.; Williams, L. D. *Nucl. Acids Res.* **2003**, *31* (5), 1536-1540.
- <sup>23</sup> Drew, H. R. Dickerson, R. E. *J. Mol. Biol.* **1981**, *151*, 535–556.
- <sup>24</sup> a) Kruger, A. T.; Lu, H.; Lee, A. H. F.; Kool, E. T. *Acc.Chem.Res.* **2007**, *40*, 141. b) Kool, E.T. *Acc.Chem.Res.* **2002**, *35*, 936.

- 
- <sup>25</sup> Lynch, S. R.; Liu, H. B.; Gao, J. M.; Kool, E. T. *J. Am. Chem. Soc.* **2006**, *128*, 14704.
- <sup>26</sup> Lee, A. H. F.; Kool, E. T. *J. Am. Chem. Soc.* **2005**, *127*, 3332.
- <sup>27</sup> Lee, A. H. F.; Kool, E. T. *J. Am. Chem. Soc.* **2006**, *128*, 2912.
- <sup>28</sup> Greco, N. J.; Tor, Y. *J. Am. Chem. Soc.* **2005**, *127*, 10784.
- <sup>29</sup> a) Li, J.-S.; Gold, B. *J. Org. Chem.* **2005**, *70*, 8764. b) Li, J.-S.; Chen, F.-X.; Shikiya, R.; Marky, L. A.; Gold, B. *J. Am. Chem. Soc.* **2005**, *127*, 12657. c) Li, J.-S.; Shikiya, R.; Marky, L. A.; Gold, B. *Biochemistry* **2004**, *43*, 1440. d) Li, J.-S.; Fan, Y.-H.; Zhang, Y.; Marky, L. A.; Gold, B. *J. Am. Chem. Soc.* **2003**, *125*, 2084.
- <sup>30</sup> Hertoghs, K. M. L.; Ellis, J. H.; Catchpole, I. R. *Nucleic Acids Res.* **2003**, *31*, 5817.
- <sup>31</sup> Kumar, N.; Nielsen, E.; Maiti, S.; Petersen, M. *J. Am. Chem. Soc.* **2006**, *128*, 14.
- <sup>32</sup> Lee, J. A.; Moon, H. R.; Kim, H. O.; Kim, K. R.; Lee, K. M.; Kim, B. T.; Hwang, K. J.; Chun, M. W.; Jacobsen, K. A.; Jeong, L. S. *J. Org. Chem.* **2005**, *70*, 5006.
- <sup>33</sup> Haraguchi, K.; Shiina, N.; Yoshimura, Y.; Shimada, H.; Hashimoto, K.; Tanaka, H. *Org. Lett.* **2004**, *6*, 2645.
- <sup>34</sup> Ober, M.; Muller, H.; Pieck, C.; Gierlich, J.; Carell, T.; *J. Am. Chem. Soc.* **2005**, *127*, 18143.
- <sup>35</sup> Wan, T. C.; Ge, Z. D.; Tampo, A.; Mio, Y.; Bienengraeber, M. W.; Tracey, W. R.; Gross, G. J.; Kwok, W. M.; Auchampach, J. A. *J. Pharmacol. Exp. Ther.* **2008**, *324*, 234-243.
- <sup>36</sup> Zhang, L. L.; Peritz, A.; Meggers, E. *J. Am. Chem. Soc.* **2005**, *127*, 4174.
- <sup>37</sup> a) Guga, P.; Maciaszek, A.; Stec, W. J. *Org. Lett.* **2005**, *7*, 3901. b) Xu, Q.; Katkevica, D.; Rozners, E. *J. Org. Chem.* **2006**, *71*, 5906.

- 
- <sup>38</sup> Nielsen, P. E.; Egholm, M. *Peptide Nucleic Acids: Protocols and Applications*. Horizon Scientific Press, Wymondham, U.K **1999**.
- <sup>39</sup> a) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991**, 254, 1497-1500. b) Egholm, M.; Buchardt, O.; Nielsen, P. E.; Berg, R. H. *J. Am. Chem. Soc.* **1992**, 114, 1895-1897. c) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Peptide nucleic acids (PNA). DNA analogues with a polyamide backbone*. In: *Antisense Research and Application*. Crook, S. and Lebleu, B. (eds.). CRC Press, Boca Raton **1993**, 363-373.
- <sup>40</sup> Nielsen, P. E.; Egholm, M.; Buchardt, O. *J. Mol. Recognit.* **1994**, 7, 165.
- <sup>41</sup> Wittung, E.; Nielsen, P. E.; Norden, B. *Biochemistry* **1997**, 36, 7973.
- <sup>42</sup> Haaima, G.; Hansen, H. E.; Christensen, L.; Dahl, O.; Nielsen, P. E. *Nucleic Acids Res* **1997**, 25, 4639.
- <sup>43</sup> Betts, L.; Josey, J. A.; Veal, J. M.; Jordan, S. R. *Science* **1995**, 270, 1838.
- <sup>44</sup> Rasmussen, H.; Kastrup, J. S.; Nielsen, J. N.; Nielsen, J. M.; Nielsen, P. E. *Nature Struct. Biol.* **1997**, 4, 98.
- <sup>45</sup> Brown, S. C.; Thomson, S. A.; Veal, J. M.; Davis, D. G. *Science* **1994**, 265, 777.
- <sup>46</sup> Eriksson, M.; Nielsen, P. E. *Nature Struct. Biol.* **1996**, 3, 410.
- <sup>47</sup> a) Nielsen, P. E. *Pure Appl. Chem.* 1998, 70, 105–110. b) Hyrup, B.; Nielsen, P. E. *Bioorg Med. Chem* **1996**, 4, 5–23.
- <sup>48</sup> Shakeel, S.; Karim, S.; Ali, A. *J. Chem. Technol. Biotechnol.* **2006**, 81, 892–899.
- <sup>49</sup> Demidov, V. V.; Frank-Kamenetskii, M. D. *Trends Biochem. Sci.* **2004**, 29, 62–71.
- <sup>50</sup> Demidov, V. V.; Potaman, V. N.; Frank-Kamenetskii, M. D.; Egholm, M.; Buchardt, O.; Sonnichsen, S. H. *et al.*, *Biochem. Pharmacol.* **1994**, 48, 1310–1313.

- 
- <sup>51</sup> Hyrup, .; Egholm, M.; Nielsen, P. E.; Wittung, P.; Norden, B.; Buchardt, O. *J. Am. Chem. Soc.* **1994**, *116*, 7964.
- <sup>52</sup> Hyrup, B.; Egholm, M.; Buchardt, O.; Nielsen, P. E. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 1083.
- <sup>53</sup> Adopted from- Krotz, A. H.; Buchardt, O.; Nielsen, P. E. *Tetrahedron Lett.* **1995**, *36*, 6941.
- <sup>54</sup> Haaime, G.; Lohse, A.; Buchardt, O.; Nielsen, P. E. *Angew. Chem.* **1996**, *35*, 1939.
- <sup>55</sup> Pless, R. C.; Ts'o, P. O.P. *Biochemistry* **1977**, *16*, 1239.
- <sup>56</sup> Miller, P. S.; Yano, J.; Yano, E.; Carrol, C.; Jayaraman, K.; Ts'o, P. O. P. *Biochemistry* **1979**, *18*, 5134.
- <sup>57</sup> a) Froehler, B.; Ng, P.; Matteucci, M. *Nucleic Acids Res.* **1988**, *16*, 4831. b) Letsinger, R. L.; Singman, C. N.; Histan, G.; Salunkhe, M. *J. Am. Chem. Soc.* **1988**, *110*, 4470.
- <sup>58</sup> Huang, Z.; Schneider, C.; Benner, S. A. *J. Org. Chem.* **1991**, *56*, 3869.
- <sup>59</sup> a) Matteucci, M. D. *Tetrahedron Lett.* **1990**, *31*, 2385. b) Veeneman, G. H.; van der Marel, G. A.; van den Elst, H.; van Boom, J. H. *Tetrahedron* **1991**, *47*, 1547; c) Vasseur, J.-J.; Debart, F.; Sanghvi, Y. S.; Cook, P. D. *J. Am. Chem. Soc.* **1992**, *114*, 4006.
- <sup>60</sup> Cheong, C.; Tinoco, I.; Chollet, A. *Nucleic Acids Res.* **1988**, *16*, 51 15.
- <sup>61</sup> Cram, D. J. *Chemtech* **1987**, *17*, 120.
- <sup>62</sup> Jencks, W. P. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1975**, *43*, 219.
- <sup>63</sup> Kool, E. T. *Chem Rev.* **1997**, *97*(5), 1473–1488.
- <sup>64</sup> a) Gunther, E. J.; Havre, P. A.; Gasparro, F. P.; Glazer, P. M. *Photochem Photobiol* **1996**, *63*, 207. b) Maher, L. *Cancer Invest* **1996**, *14*, 66. c) Raha, M.; Wang, G.; Seidman, M. M.; Glazer, P. M. *Proc Natl Acad Sci USA* **1996**, *93*, 2941.

- 
- <sup>65</sup> Afonina, I.; Kutuyavin, I.; Lukhtanov, E.; Meyer, R. B.; Gamper H. *Proc Natl Acad Sci USA* **1996** *93*, 3199.
- <sup>66</sup> Morishita, R.; Gibbons, G. H.; Horiuchi, M.; Ellison, K. E.; Nakama, M.; Zhang, L.; Kaneda, Y.; Ogihara, T.; Dzau, V. J. *Proc Natl Acad Sci USA* **1995**, *92*, 5855.
- <sup>67</sup> Kurreck, J. *Eur. J. Biochem.* **2003**, *270*, 1628.
- <sup>68</sup> a) Crooke, S. T. *Therapeutic Applications of Oligonucleotides*, Springer, Heidelberg, **1995**. b) Stein, C. A.; Krieg, A. M. *Applied Antisense Oligonucleotide Technology*, John Wiley & Sons, New York, **1998**. c) Crooke S. T. *Handbook of Experimental Pharmacology: Antisense Research and Applications*, Vol. 131, Springer, Berlin, Heidelberg, **1998**.
- <sup>69</sup> a) Jacobsen, N.; Bentzen, J.; Meldgaard, M.; Jakobsen, M. H.; Fenger, M.; Kauppinen, S.; Skouv, J. *Nucleic Acids Res.* **2002**, *30*, e100. b) Jacobsen, N.; Fenger, M.; Bentzen, J.; Rasmussen, S. L.; Jakobsen, M. H.; Fenstholt, J.; Skouv, J. *Clin. Chem.* **2002**, *48*, 657. c) Orum, H.; Jakobsen, M. H.; Koch, T.; Vuust, J.; Borre, M. B. *Clin. Chem.* **1999**, *45*, 1898. d) Lattora, D.; Hopkins, D.; Campbell, K.; Hurley, J. M. *BioTechniques* **2003**, *34*, 1154. e) Thomassin, H.; Kress, C.; Grange, T. *Nucleic Acids Res.* **2004**, *32*, e168.
- <sup>70</sup> Kaur, H.; Wengel, J.; Maiti, S. *Biochem. Biophys. Res. Commun.* **2007**, *352*, 118.
- <sup>71</sup> Valoczi, A.; Hornyik, C.; Varga, N.; Burgyan, J.; Kauppinen, S.; Havelda, Z. *Nucleic Acids Res.* **2004**, *32*, e175.
- <sup>72</sup> Veselkov, A. G.; Demidov, V. V.; Neilsen, P. E.; Frank-Kamenetskii, M. D. *Nucl. Acids Res.* **1996**, *24*, 2483-2488.
- <sup>73</sup> Demidov, V. V. *Expert Rev Mol Diagn* **2001**, *1*, 343–351.



- 
- <sup>74</sup> Ørum, H. *Curr. Issues Mol. Biol.* **2000**, *2(1)*, 27-30.
- <sup>75</sup> Ørum, H.; Nielsen, P. E.; Egholm, M.; Berg, R.H.; Buchardt, O.; Stanley, C. *Nucl. Acid Res.* **1993**, *21*, 5332-5336.
- <sup>76</sup> Demidov, V. V.; Bukanov, N. O.; Frank-Kamenetskii, M. D. *Curr. Issues Mol. Biol.* **2000**, *2(1)*, 31-35.
- <sup>77</sup> Else, M.; Dearden, C. E.; Matutes, E., et al. *Br. J. Haematol.* **2009**, *145* (6), 733–740.
- <sup>78</sup> a) G. Acs, G.; Reich, E.; Mori, M. *Proc. Natl. Acad. Sci. U.S.A.* **1964**, *52*, 493-501. b) el Kouni, M.H.; Diop, D.; Cha, S. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 6667-6670. c) Fabianowska-Majewska, K.; Duley, J. A.; Simmonds, H. A. *Biochem. Pharmacol.* **1994**, *48*, 897-903. d) Stahmann, K.P., et al. *Appl. Microbiol. Biotechnol.* **2000**, *53*, 509-516.
- <sup>79</sup> a) Pugh, C. S.; et al. *J. Biol. Chem.* **1978**, *253*, 4075-4077. b) Zingg, J.M., et al. *Nucleic Acids Res.* **1996**, *24*, 3267-3275. c) Lin, Q., et al. *J. Am. Chem. Soc.* **2001**, *123*, 11608-11613. d) Cheng, D., et al. *J. Biol. Chem.* **2004**, *279*, 23892-23899.
- <sup>80</sup> a) Schneider-Stock, R. et al *J.Pharmacol.Exp.Ther.* **2005**, *312*, 525. b) Mikkelsen, T. S. et al *Nature* **2008**, *454*, 49.
- <sup>81</sup> a) Zhong, M.; Nowak, I.; Cannon, J. F.; Robins, M. J. *J. Org. Chem.* **2006**, *71*, 4216–4221. b) Zhong, M.; Nowak, I.; Robins, M. J. *J. Org. Chem.* **2006**, *71*, 7773–7779.
- <sup>82</sup> Arico, J. W.; Calhoun, A. K.; Salandria, K. J.; McLaughlin, L. W. *Org. Lett.* **2010**, *12*, 120-122.
- <sup>83</sup> a) Koshkin, A. A.; Nielson, P.; Meldgaard, M.; Rajwanshi, V. K.; Singh, S. K.; Wengel, J. *J. Am Chem. Soc.* **1998**, *120*, 13252. b) Singh, S. K.; Wengel, J. *Chem. Commun.* **1998**, *12*, 1247. c) Wengel, J. *Acc. Chem. Res.* **1999**, *32*, 301.
- <sup>84</sup> Branda, N.; Kurz, G.; Lehn, J. M. *Chemical Communications* **1996**, 2443-2444.

- 
- <sup>85</sup> Branda, N.; Kurz, G.; Lehn, J-M *Chem. Commun.* **1996**, 2443-2444.
- <sup>86</sup> Chen, D.; Meena; Sharma, S. K.; McLaughlin, L. W. *J. Am. Chem. Soc.*, **2004**, *126*, 70-71.
- <sup>87</sup> Chen, H.; Meena, McLaughlin, L. W. *J. Am. Chem. Soc.*, **2008**, *130*, 13190–13191.

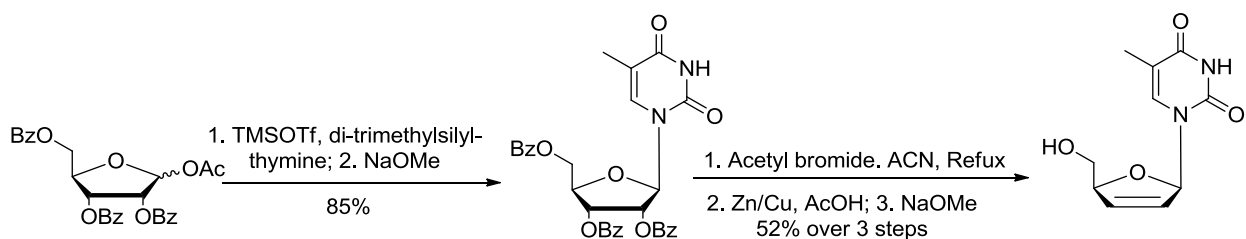
## **Chapter 2**

# **Bicyclo-1,1'-dicarboxylic Acid Cyclic-anhydride as a Directing/Protecting Group Purine Glycosylations**

## Background

The advancement of automated, solid-phase synthetic procedures for the synthesis of DNA<sup>1</sup> and RNA<sup>2</sup> has in the last couple of decades led to widespread use of natural and modified nucleic acid analogues in chemistry, medicine, biology and many interdisciplinary fields. While synthesized DNA strands having the natural structure are most widely used as primers, linkers, and probes in molecular biological techniques, there are a host of possible modified DNA structures with many new properties which offer a host of new possibilities. One of those structures can be achieved by incorporating modified nucleoside monomers in the DNA or RNA strands; with modified base, modified sugar moieties, or modification involving both base and sugar.

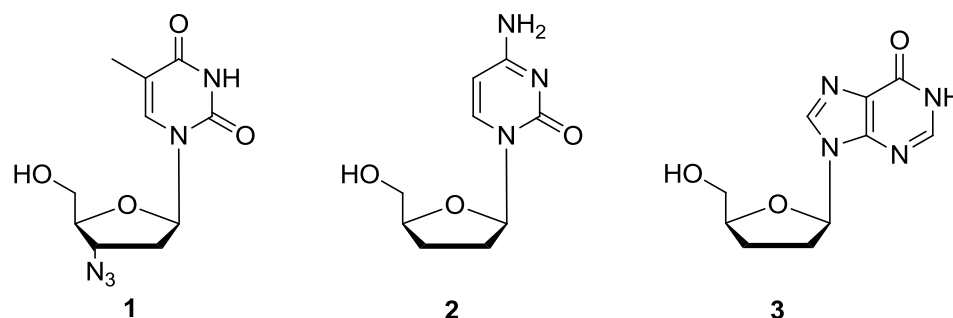
In literature, it has been well documented that modified nucleoside monomers can be chemically synthesized via divergent or convergent approaches or combination of both (Scheme 2.1). The example shown illustrates this concept. Step 1 is an example of convergent methodology whereas the subsequent steps show the divergent procedure<sup>3</sup>.



**Scheme 2.1:** Example of a combination of convergent and divergent approaches<sup>3</sup>.

The divergent approach<sup>4</sup> became popular when the modification of intact nucleosides was the first method available for synthesizing 2',3'-dideoxynucleosides. In fact, the

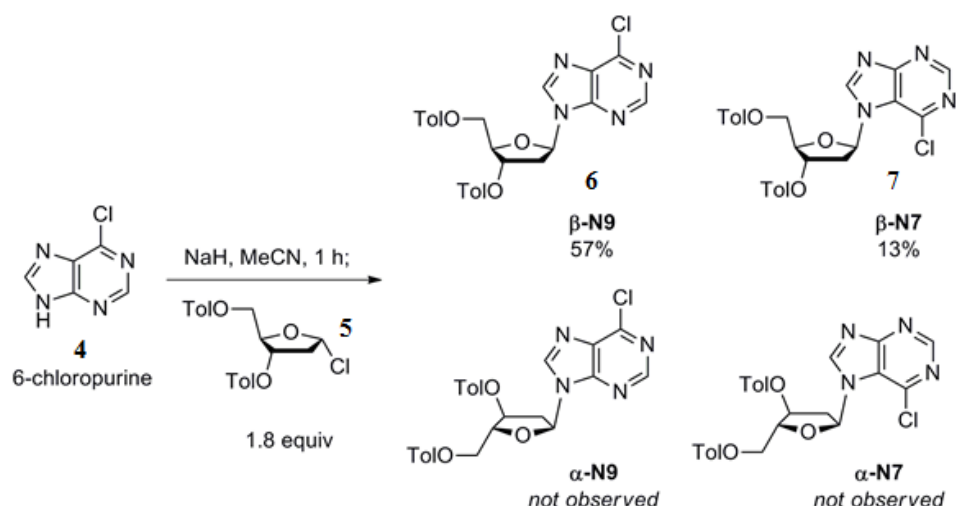
compounds currently approved by the FDA for the treatment of AIDS (AZT **1**, ddC **2**, ddI **3**; Figure 2.1) were first synthesized using this approach. These routes rely largely, but not exclusively, on strategies which involve the deoxygenation of furanonucleosides. The main advantage of these approaches is that control of the glycosidic stereochemistry is not an issue, i.e., it is either set by nature or can be controlled synthetically via anchimeric assistance of an appropriate 2'-substituent. However, they are limited primarily by the availability of naturally occurring nucleoside starting materials (e. g., thymidine, cytidine, uridine, adenosine, guanosine).



**Figure 2.1:** Structure of : 1. AZT; 2. ddC, 3. Ddl.

Alternately, convergent approaches<sup>4</sup> to synthesize nucleoside derivatives are more versatile than other approaches, since they involve the coupling of a carbohydrate moiety (unmodified or modified) with a purine or pyrimidine base (unmodified or modified) via a nitrogen glycosylation reaction. Since these approaches enable one to change both the base and carbohydrate components of derivative monomers, they represent the preferred strategy for preparing substrates for nucleic acid incorporation and their subsequent biochemical and bio-physical studies. In practice, however, this only holds true when the glycosylation reaction proceeds with a reasonable degree of stereochemical control. The

challenges are apparent in the synthesis of purine 2'-deoxynucleosides, because (1) there are theoretically four possible products (Scheme 2.2), and (2) the relative ratios are more substrate-dependent. There are several methods for purine glycosylation, and each affords good to excellent selectivity for some substrates, but a general method applicable to the synthesis of all purine 2'-deoxyribonucleosides remains elusive. Certainly the most common method employed with simple heterocycles such as 6-chloropurine **4** is the sodium salt method<sup>5</sup>; the most common sugar employed is 2-deoxy-3,5-di-*O*-(*p*-toluoyl)- $\alpha$ -D-*erythro*-pentofuranosyl chloride **5**<sup>6</sup>. The glycosylation of **4** gives the desired **6** in 57% yield with 13%  $\beta$ -N7 product **7** and none of the possible  $\alpha$ -nucleoside products (Scheme 2.2).

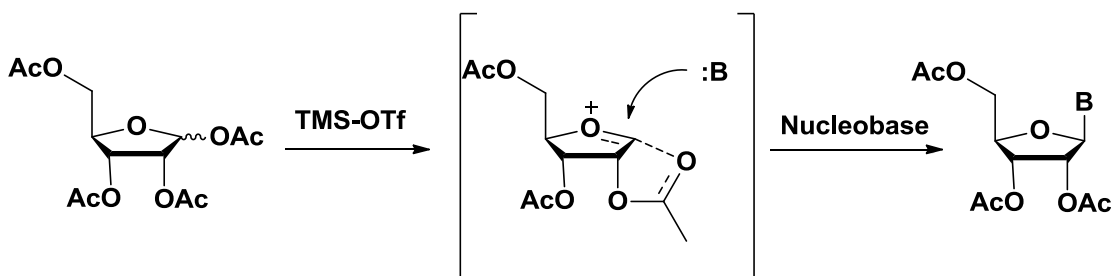


**Scheme 2.2.** The four possible purine coupling products for the reaction of 6-chloropurine with 2-deoxy-3,5-di-*O*-(*p*-toluoyl)- $\beta$ -D-*erythro*-pentofuranosyl chloride.

### 2.1. Common Routes to Prepare 2'-deoxyribonucleosides

For modified ribonucleosides, the issue of  $\alpha$ -anomers can be reasonably avoided by the neighboring group effect of groups like benzoyl or acetyl groups at the 2'-position

(Scheme 2.3)<sup>7</sup>. It is possible to utilize these ribonucleosides to access modified 2'-deoxyribonucleosides by removing the 2'-hydroxyl by methodologies like Barton deoxygenation.

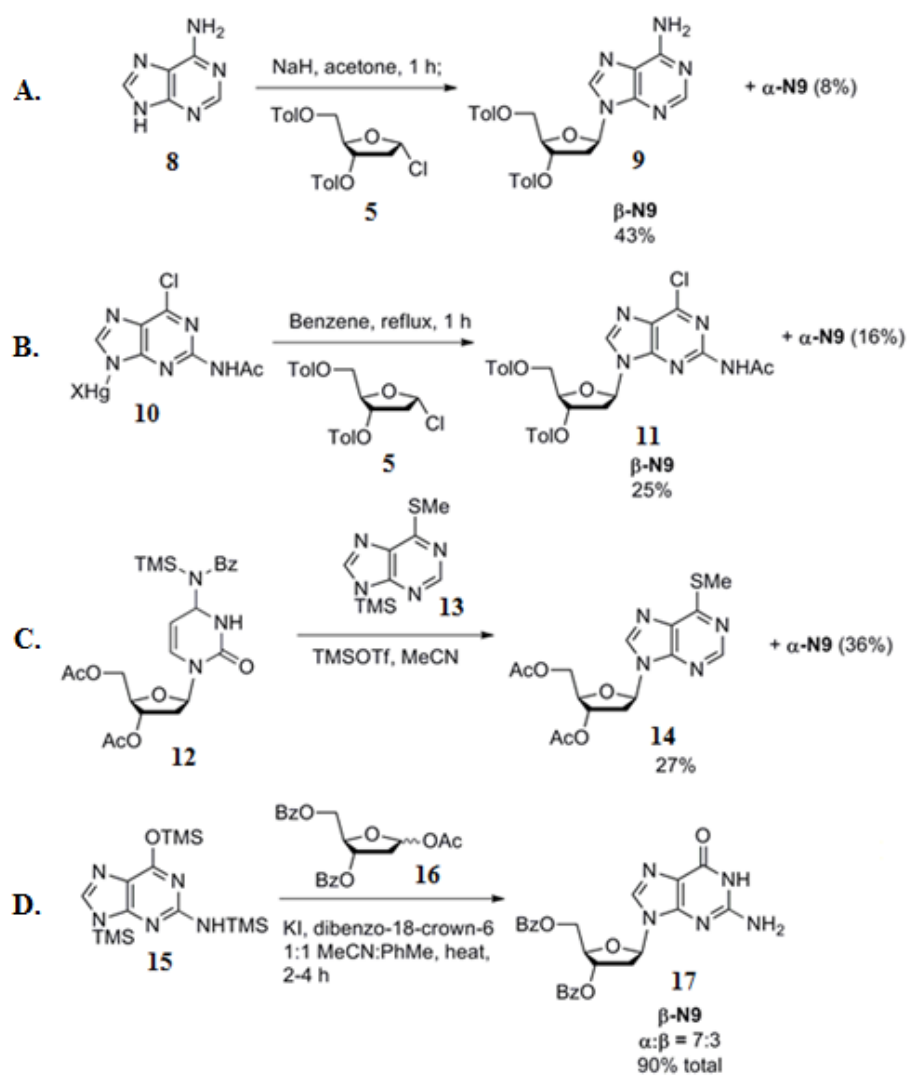


**Scheme 2.3:** Neighboring group effect operating in ribonucleoside synthesis.

Apart from this approach, there are several procedures for the direct access of 2'-deoxyribonucleosides. Some of them have been summarized below and illustrated in Scheme 2.4.

- a) Sodium salt method (Figure 2.4 A): Adenine (**8**) can be converted to its sodium salt with NaH in acetone, which reacts with **5** to form protected 2'-deoxyadenosine (**9**) in 48% yield with 8%  $\alpha$ -N9 product<sup>8</sup>. It was observed that the sodium salt of **8** has poor solubility in non-polar solvents like DCM or dichloroethane resulting in a slow reaction rate, thereby increasing the anomerization of **5** and subsequently producing higher amounts of the  $\alpha$ -N9 product. When the reactions were carried out in polar solvents like ACN or DMF, the sodium salt of the bases were found to react faster. However, polar solvents also facilitate the anomerization of sugar as well, resulting in more  $\alpha$ -nucleoside product.

b) Metal salt method (Figure 2.4 B): One of the first glycosylation methods developed for nucleosides was the metal salt procedure in which the purine is converted to a heavy metal salt, typically Hg, and reacted directly with the sugar. Employing this method compound **10** yielded the desired product **11** in a 25% yield with 16%  $\alpha$ -N9 product<sup>9</sup>, a remarkable decrease from the 57% yield obtained with **4**. A serious drawback of this method is the removal of the highly toxic heavy metals like Hg after the reaction.



**Scheme 2.4:** A few approaches to access 2'-deoxyribonucleosides via different glycosylation methods.



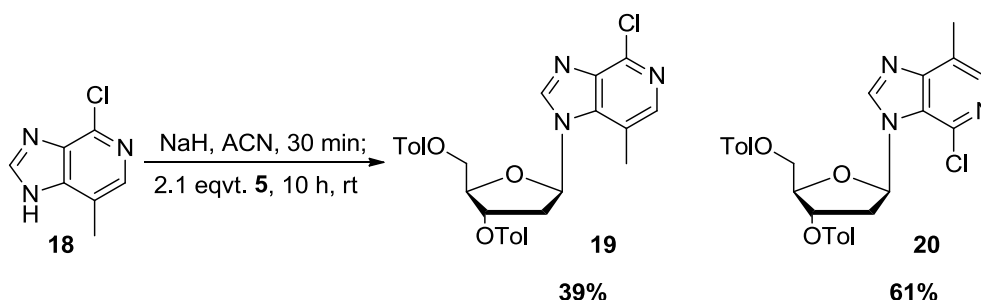
- c) Transglycosylation method (Figure 2.4 C): This method (**12**) takes advantage of the fact that nucleoside formation in the presence of a Lewis acid is a reversible process. Lewis acids such as  $\text{HgBr}_2$ ,  $\text{TMSClO}_4$  or  $\text{TMSOTf}$  are effective in promoting this type of glycosylation. In the case of 2-deoxyribose sugars significant amounts of the  $\alpha$ -nucleoside product are usually obtained. **12** reacts with the silylated nucleobase **13** to afford 27% **14** and 36% of the corresponding  $\alpha$ -N9 product<sup>10</sup>. When **12** reacts with silylated nucleobase **13** in presence of  $\text{TMSOTf}$ , 27% of the desired **14** and 36% of the corresponding  $\alpha$ -N9 product were reported<sup>10</sup>.
- d) Phase Transfer Method (Figure 2.4 D): Alkali halides have been used to catalyze glycosylation reaction. Silylated  $\text{N}^2$ -acetyl guanine **15** reacts with **16** in the presence of KI and a crown ether in 1:1 MeCN:PhMe yielding **16** in 90% as a 7:3 mixture of  $\alpha$ : $\beta$  anomers<sup>11</sup>.

Even though the sodium salt method for glycosylation reaction has issues with regio- and stereo-selectivity, it has been accepted as the most effective methodology to perform glycosylation reactions for this substrate class, due to better results and ease of performing the procedure over other methods. However, it has been found that sodium salt is affected by the duration of the reaction, solvent, temperature and counter-ion. Investigation has proven that longer reaction time and polar solvents increase the relative ratio of  $\alpha$ : $\beta$ . The reason being that these factors individually or in some combination favor the anomerization of the sugar **5**<sup>12</sup>. More  $\beta$ -nucleoside would form initially, but could equilibrate over time to form more  $\alpha$  product. Apart from these external factors, the inherent factors such as sterics and electronics of the heterocycle might influence the

results of the reaction as well. Considering both sterics and electronics of the nucleobase favor glycosylation at N9, the efficiency of coupling becomes a question of the rate of glycosylation vs. the rate of the anomerization of **5**.

## 2.2. Problems with Traditional Routes for our Modified Nucleoside Synthesis

In terms of inherent factors, the sodium salt method works reasonably well for electron-deficient nucleobase like chloropurine and even for aminopurines like adenine. But increasing electron density and steric bulk at the 3-position of 3-deaza-3-substituted purines affect the efficiency of glycosylation adversely. For minor groove modifications, we need to install a sterically demanding substituent in the 3-position of 3-deaza-3-substituted purines. Some of these substituents increase the electron density of the heterocycle. When such substituents were used to couple with the sugar, **5**, the regioselectivity was affected. For example, when 3-deaza-3-methyl-6-chloropurine **18** was coupled to **5**, only 39% of the desired  $\beta$ -N9 product **19** was obtained<sup>13</sup>, a reversal of regioselectivity (Scheme 2.5). Similar trends were observed with other 3-deaza-3-substituted purines. This became a stumbling block for material throughput in our efforts to synthesize modified nucleosides required for minor groove modification. Large-scale glycosylations of **18** typically resulted in yields of 25-30% of **19**, which was also quite difficult to separate from the N7 product **20**. For our purposes we need to install the amine group at the 6-position. This was found to be unprecedentedly difficult. The amination of **19** required a harsh condition of overnight treatment with hydrazine followed by a 6 h reaction with Raney nickel that only resulted in a 50-60% yield.



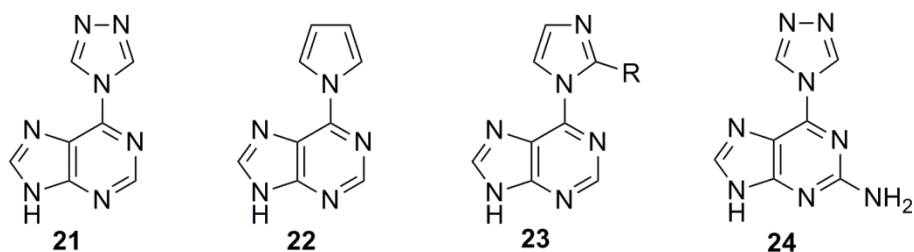
**Scheme 2.5:** Reverse regioselectivity with **18**.

Further investigation showed that more of the desired N9 product consistently was obtained with increasing the size of the substituent at the 6-position. Our hypothesis is that a directing group, which can also have protecting group functionality, with sufficient steric bulk placed in the 6-position would increase the relative yield of the desired N9 product. This should offer a significant improvement over the current methodology, for both modified purines like 3-deaza-3-substituted purines and unmodified purines. The use of such a group in the 6-position could also increase the solubility of the modified purine, thereby speeding up the reaction, which in turn could increase the yield of the glycosylation. Such an improvement would be a significant contribution to the field of nucleoside synthesis, since often it is necessary to settle for a glycosylation yield that is rather poor when considered in the overall context of organic synthesis, due to the lack of alternatives for difficult couplings.

### 2.3. Development of Bulky Substituent at C6-position of Purines

Various research groups have pursued the concept of a directing group. Bulky groups in the 6-position of purines are known to influence the N9:N7 ratios of alkylations<sup>14</sup> and

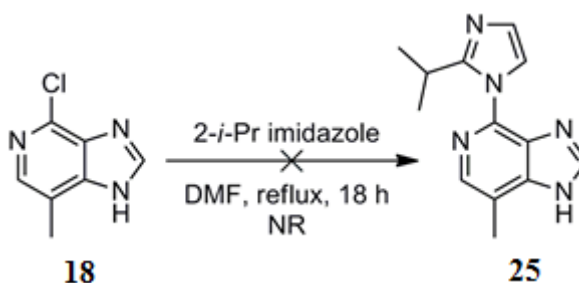
glycosylations<sup>15</sup>. Robins<sup>16</sup>, Mintas<sup>17</sup>, Demeunynck<sup>18</sup> and others have worked in this area to develop different bulky functionality at the 6-position. Robins' work involved 6-(1,2,4-triazol-4-yl)purine **21** (**Figure 2.2**), 6-(2-substituted-imidazol-1-yl)purine **23** (R= H, Pr, Bu, Ph etc., for tuning the solubility of the purine) and 6-(imidazol-1-yl)purine. Mintas and Demeunynck investigated 6-pyrrole-substituted purines **22**. Demeunynck also worked with 6-(1,2,4-triazol-4-yl)-2,6-diaminopurine. When **21** was alkylated, they yielded the N9 product exclusively<sup>19</sup>. However, the application of these groups for glycosylation reactions had serious problem as these purines are very polar and required solvent like DMF for the glycosylation reaction. The results with **22** were not promising either. It was found that it is an inefficient directing group and removal of the group was a challenge. The lipophilicity of **23** can be tuned to be suitable for glycosylation reaction. It afforded N9 products exclusively, but the  $\alpha:\beta$  ratio was unfavorable.



**Figure 2.2:** Selected examples of purines with different groups at 6-position of the purine.

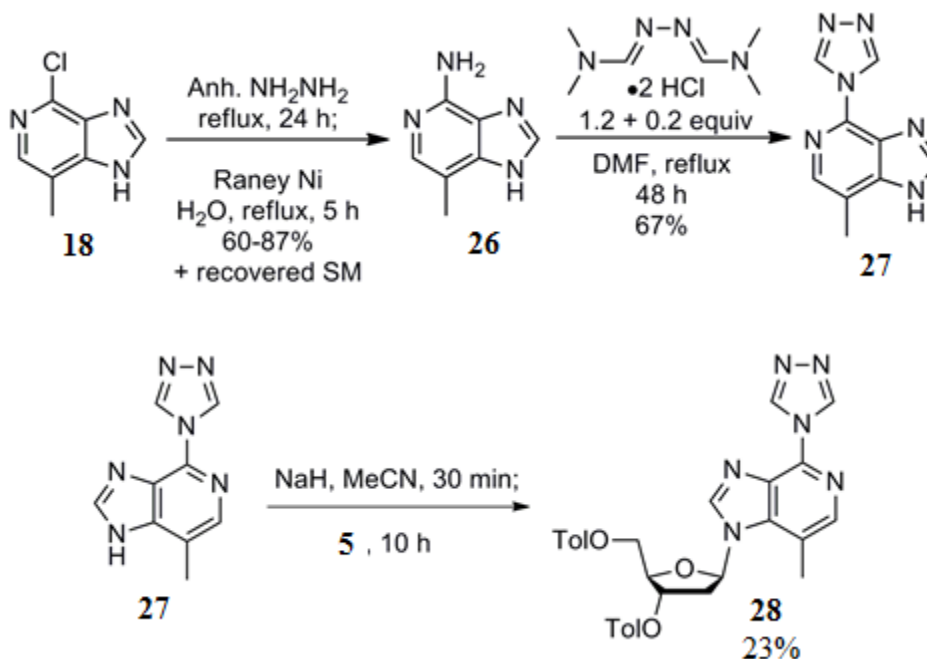
Some of these methodologies were investigated in our laboratory. Unfortunately they proved not suitable for glycosylation reactions involving 3-deaza-3-methyl-6-chloro purine **18**. Some of results are discussed below.

- a) 6-chloropurines was found to react with imidazoles at 65 °C in DMF in significant yield. However when **18** reacted with 2-isopropylimidazole, no product was detected (Scheme 2.6). The reactivity was found to be so poor that even after refluxing for 18 h, no appreciable amount of **25** was observed.



**Scheme 2.6.** Failure to install 2-isopropylimidazole.

- b) In another attempt, a triazole was installed at the C6 position of **18**. To do that, the 6-Cl was first converted to an amino group by reacting **18** with anhydrous hydrazine at reflux for 24 h, then reduced with Raney nickel in H<sub>2</sub>O to obtain 60–87% of **26**. The yield of this reaction was variable, with 10–30% starting material typically recovered. Azine reagent did react with **26** to afford the 6-(1,2,4-triazol-4-yl)-3-deaza-3-methylpurine **27**. Reaction of this extremely insoluble compound with NaH in MeCN followed by **5** resulted in a N9-selective glycosylation, but **28** was only isolated in 23% yield (Scheme 2.7).



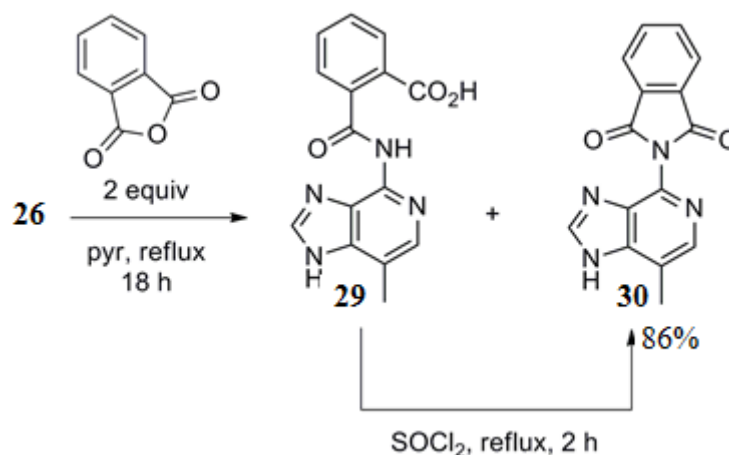
**Scheme 2.7:** Attempt to use triazole as protecting/directing group.

- c) Further, deprotection of the imidazole blocking group was difficult<sup>20</sup>. For 6-(2-substituted-imidazol-1-yl)-2-chloro-2'-deoxyadenosine nucleosides treatment with  $\text{NH}_3/\text{MeOH}$  at  $80^\circ\text{C}$  for 15 h in a pressure vessel gave 2-chloro-2'-deoxyadenosine but separation of the imidazole from the nucleoside was problematic. For purines lacking the 2-chloro substituent, ammonolysis was more difficult. A selective deprotection of toluyl groups while keeping the protecting/directing group was necessary for our 3-deaza-3-substituted purines.

#### 2.4. Screening New Bulky Substituent at C6-position of Purines

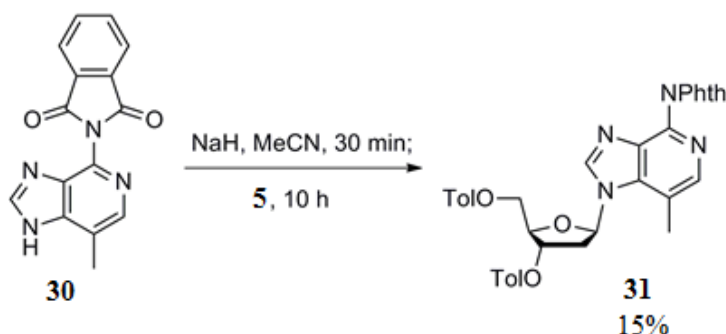
Since these groups failed to serve our purposes, the next choice was to try different cyclic anhydride. Initial efforts were made using phthaloyl group and *cis*-1,4-cyclohexanedicarboxylic anhydride. Brief discussion about their success and limitations are described below.

a) Using phthaloyl group: The obvious choice of this group lies in the fact that it is commercially available. **26** was allowed to react with phthalic anhydride under refluxing conditions in anhydrous pyridine to yield a mixture of desired **29** and the incomplete diimide **30**. Various separation methods including column chromatography and recrystallization could not separate **29** and **30**. Instead refluxing the product mixture in excess  $\text{SOCl}_2$  for 2 h completely transformed the amide into **30** via the conversion of the acid of **29** into the acid chloride, which is highly reactive and cyclizes quickly to give **30** (Scheme 2.8).



**Scheme 2.8:** Installing phthaloyl group on **26**.

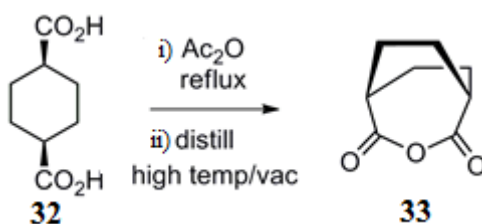
When **30** reacted with **5**, using the sodium salt method, for glycosylation reaction, only 15% of the desired  $\beta$ -N9 was observed. The overall yield of the reaction was 65% but the remainder of the product was a mixture of other regio- and stereoisomers ( $\alpha$ -N9,  $\alpha$ -N7 and  $\beta$ -N7) (Scheme 2.9). Separation of these isomers was a challenge.



**Scheme 2.9:** Glycosylation reaction of **30**.

Although it is reported that phthaloyl groups need hydrazine to be removed from aliphatic amines, most of the substrates we worked with were found to be labile under mild basic conditions. This makes it difficult to remove toluyl groups while keeping the phthaloyl group in place for phosphoramidite chemistry and subsequent incorporation in DNA. The amine can be re-protected but this adds more steps to get the phosphoramidite and, hence, is not an ideal solution.

- b) Using *cis*-1,4-cyclohexanedicarboxylic anhydride<sup>21</sup>: This was obtained by refluxing **32** in acetic anhydride for 5 h, followed by distillation (Scheme 2.10). There were two major road-blocks for using this group: i) it tends to form polyanhydrides fairly easily and, ii) the installation on simple adenine did not work.

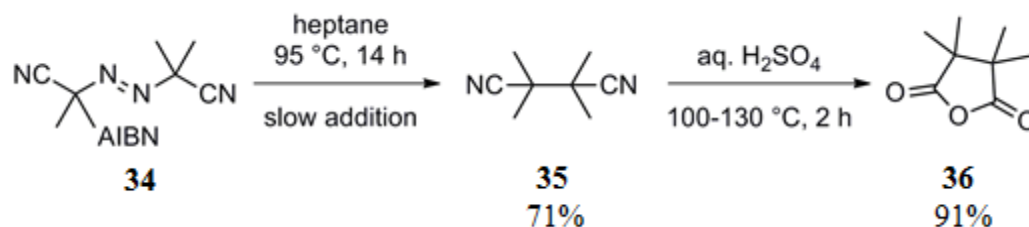


**Scheme 2.10:** Synthesis of *cis*-1,4-cyclohexanedicarboxylic anhydride.



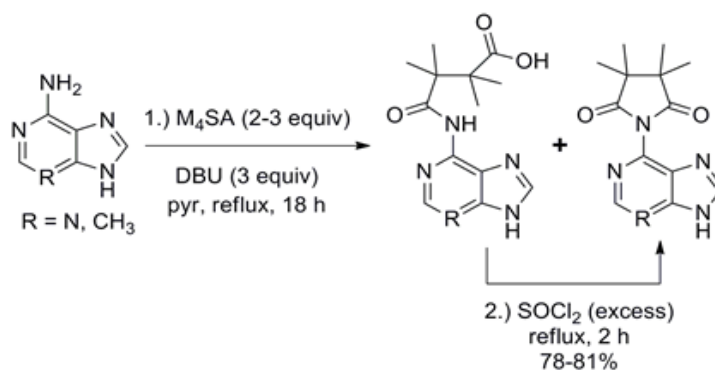
As a result this did not proved to be a good candidate for our purposes either.

- c) Using tetramethylsuccinic anhydride (M<sub>4</sub>SA)<sup>22</sup>: The use of M<sub>4</sub>SA as a protecting/directing group for purine glycosylation to achieve desired  $\beta$ -N9 product resulted in a significant improvement compared to known methodologies. Slow addition of AIBN over 14 h to hot heptane resulted in homocoupling of the radical fragments to afford dinitrile **35**. It was suspended in aqueous H<sub>2</sub>SO<sub>4</sub> and heated for 2 h with vigorous stirring to obtain the anhydride in excellent yield (Scheme 2.11).



**Scheme 2.11:** Synthesis of M<sub>4</sub>SA.

**36** was used to protect the C6-amino group of adenine and **26**. This resulted in a mixture of desired dimide and mono-imide products. The mono-imide was converted to diimide by refluxing in excess thionyl chloride for 2 h. The total yield was in the range of 78-81% (Scheme 2.12).



**Scheme 2.12:** Installing M<sub>4</sub>SA at C6 of adenine and **26**.

The diimide was used for the glycosylation reaction via the sodium salt method. This proved to be a significant improvement over current methodology. These results will be discussed in detail and compared with other groups in section 2.6. The crystal structure of the M<sub>4</sub>SI of adenine was obtained and this provided valuable insights about why M<sub>4</sub>SI acts as an effective directing group for the glycosylation reaction. In the "face view" (Figure 2.3a) the four methyl groups can be observed positioned "umbrella-like" above the N7 edge of the adenine ring system. The distance between the imide carbonyl oxygen and the N7 nitrogen is 3.1 Å, and that to C5 is 3.0 Å. These distances likely prevent a coplanar conformation between the imide and heterocycle, and in combination with the four methyl groups seem sufficient to limit access to the N7 by the incoming sugar. The "edge view" (Figure 2.3b) illustrates that the plane of the imide functionality is in fact rotated 52° relative to the plane of the heterocycle, and this rotation is necessary to prevent the steric clash between the imide carbonyls and the N7 nitrogen and C5 carbon.

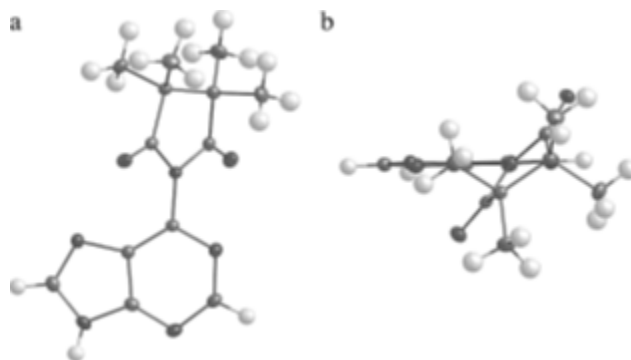
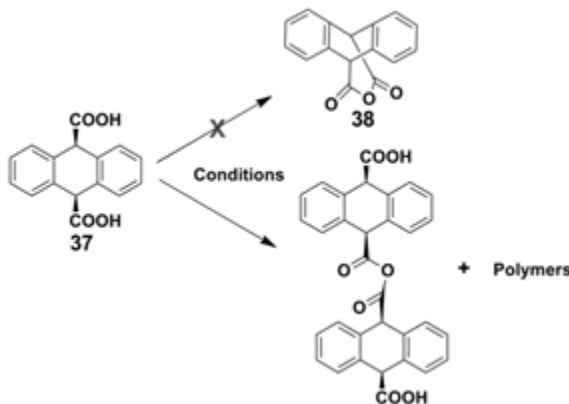


Figure 2.3: Crystal structure of adenyln-N<sup>6</sup>-tetramethylsuccinimide. (a) face view, (b) edge view

With these findings, we decided to investigate further whether increasing the steric bulk at the sites of the methyl groups would optimize the directing ability of substituted succinimide derivatives. It led us to explore the following two groups.

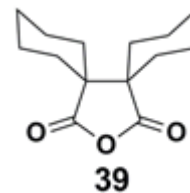
- d) Using 9,10-dihydroanthracene-9,10-dicarboxylic acid cyclic-anhydride: If steric bulk was the reason for the efficiency of M<sub>4</sub>SA then two aromatic rings would be an ideal candidate to test our hypothesis. To synthesize 9,10-dihydroanthracene-9,10-dicarboxylic cyclic-anhydride (**38**), we started from commercially available *cis*-9,10-dihydroanthracene-9,10-dicarboxylic acid (**37**). A variety of conditions were tried, but we were unable to isolate the desired anhydride **38**. In most of the cases we ended up getting either dimer or polymers (Scheme 2.13).



**Scheme 2.13:** Attempts to synthesize **38**.

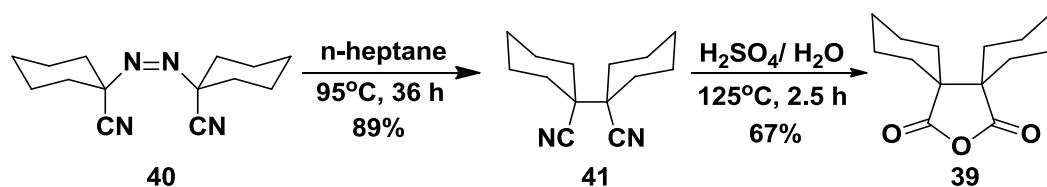
Another route to try was to directly install **37** at the C6-amino of adenine or **26**. Unfortunately, we were never able to obtain the desired product. Since acid chlorides are more reactive than anhydrides, the corresponding acid chloride were prepared *in situ* and allowed to react with adenine. No product was isolated. After attempting different procedures to perform this reaction, it was decided to try the next anhydride.

e) Using bicyclo-1,1'-dicarboxylic acid cyclic-anhydride (Cy<sub>2</sub>SA) (**39**): This group was found to be a highly efficient protecting/directing group. In the next section, we are going to discuss the synthesis and applicability of Cy<sub>2</sub>SA group as a protecting/directing group for purine glycosylation and compare its result with some of the other groups, especially M<sub>4</sub>SA group.



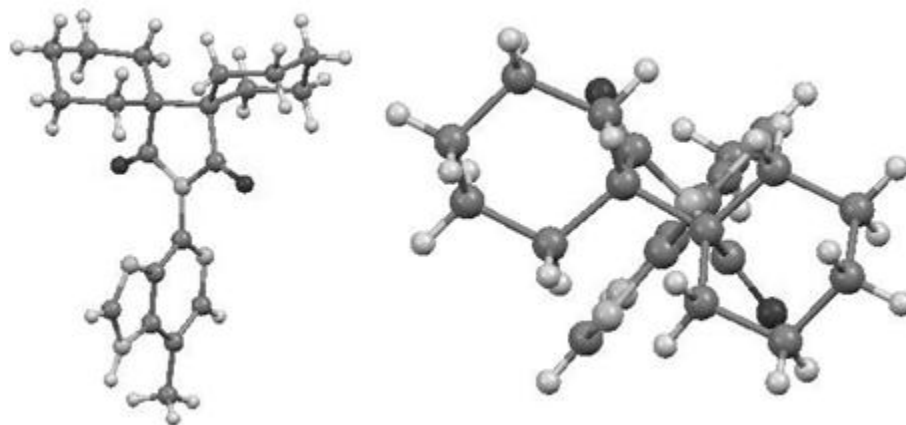
## 2.5. Bicyclo-1,1'-dicarboxylic Acid Cyclic-anhydride as a Directing/Protecting Group

The synthesis of Cy<sub>2</sub>SA is presented in Scheme 2.14. We could easily convert 1,1'-azobis(cyclohexanecarbonitrile) (**40**) to the corresponding bicyclohexyl-1,1'-dinitrile (**41**) (Scheme 2.14) and then subject the product to strong acidic conditions<sup>23</sup> to form the corresponding cyclic anhydride (**39**) (Scheme 2.14). The acidic cyclization step was tricky. During the first few attempts, only 12-15% product was isolated. The reaction was messy and it was difficult to monitor the reaction. A lot of side-products were obtained including incomplete hydrolyzed products, di-acid derivative etc. It was virtually impossible to separate them. After optimizing the conditions, it was found that the mixture of conc. sulfuric acid and water in the ratio of 2:1 was the optimum to carry out the reaction. The reaction mixture was heated at around 110 °C (not higher than that) for 1.5 hr followed by heating at 125 °C for another 1.5 hr.



**Scheme 2.14.** Preparation of the Bicyclo-1,1'-dicarboxylic Acid Cyclic-anhydride (Cy<sub>2</sub>SA) group.

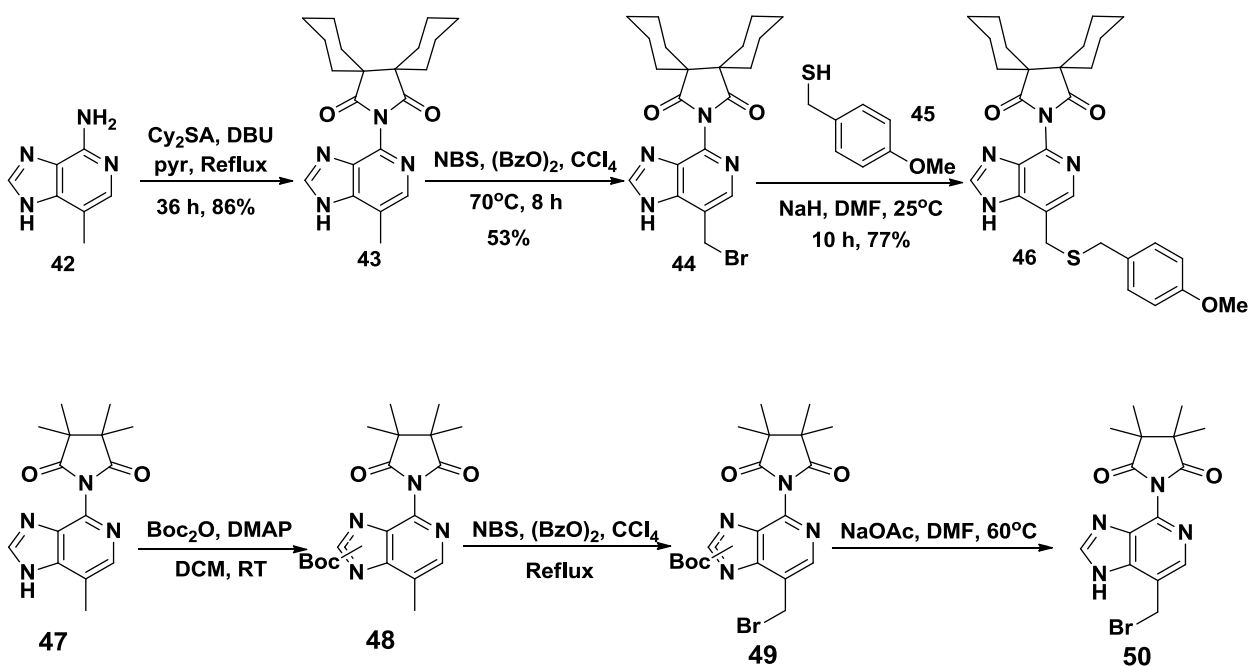
Introduction of the Cy<sub>2</sub>SI protecting group to a selection of 6-amino purines is accomplished in a single step by simply refluxing the two reactants in the presence of DBU in pyridine (no thionyl chloride is necessary to effect the cyclization as in the case of M<sub>4</sub>SI installation<sup>22</sup>). The diimide product is obtained in reasonable yields. The crystal structure of the 3-methylated 3-deaz-4-methyladenine protected derivative illustrates (Figure 2.4) that the two cyclohexyl groups are crowding over the N7 of the purine base; specifically one methylene group from each of the cyclohexyl moieties are spatially positioned directly above the N7 (top view, Figure 2.4).



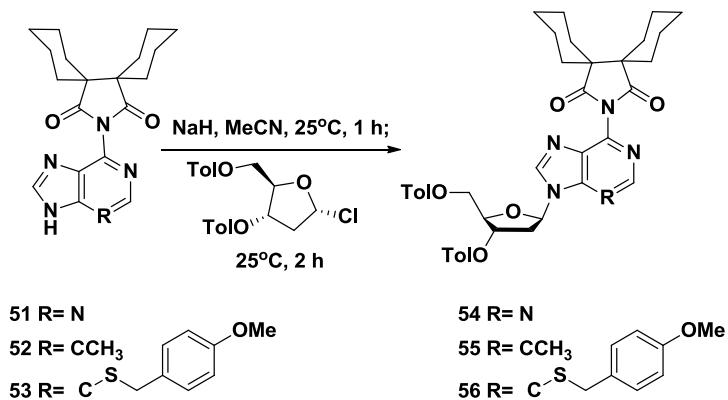
**Figure 2.4.** Crystal structure of the 3-deaza-3-methyl-adenine containing C6-amino protected with Cy<sub>2</sub>SA diimide from X-ray diffraction methods: a) side view and b) top view.

Glycosylation by the sodium salt method of unmodified or modified purine derivatives with C6-amino group is extremely effective with the Cy<sub>2</sub>SI protecting group in two aspects: (i) the –NH<sub>2</sub> functionality, likely to react with the sodium hydride activation agent is absent and replaced by the much less reactive cyclicdiimide (Scheme 2.16), and (ii) the steric bulk of Cy<sub>2</sub>SI significantly blocks glycosylation at N7 resulting in higher yields of the desirable N9 product (Scheme 2.17, Figure 2.4).

To investigate the efficiency of the Cy<sub>2</sub>SI group as directing group, we needed appropriate nucleobases. Adenine was used as the control base. We prepared two other bases (**43** and **44**) from **42** (Scheme 2.15 Top). One of the advantages of this new group, compared to M<sub>4</sub>SI, is that it imparts non-polarity to the molecule. As a result the conversion of **44** to **46** can be performed in one step due to the partial solubility of **44** in carbon tetrachloride (Scheme 2.15 Top). For the similar conversion with M<sub>4</sub>SI analogue (Scheme 2.16 Bottom), **47** needs to be first Boc-protected. That increases the non-polarity of the molecule **48** which is then used to install the bromide to prepare **49**. Then the Boc-group needs to be removed to get **50**. Hence, when using Cy<sub>2</sub>SI in this specific conversion, the number of steps can be reduced and the overall yield can be increased. Once we have the bromide (**44**), this can be replaced by a variety of nucleophiles to introduce different minor groove modification functionality. For our purposes we decided to substitute the bromide of **44** by (4-methoxyphenyl)methanethiol (**45**) to prepare **46**.



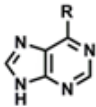
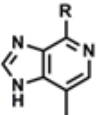
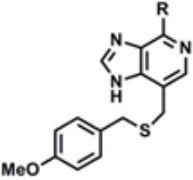
**Scheme 2.15.** Top- Preparation of modified nucleobases. Bottom- Three-step process of installing bromide functionality using M<sub>4</sub>SI.



**Scheme 2.17.** Glycosylation of 3-substituted purines using bicyclo-1,1'-dicarboxylic acid cyclic diimide protected amino group.

The results of the glycosylation results have been tabulated in Table 2.1. With the simple adenine heterocycle both M<sub>4</sub>SI and Cy<sub>2</sub>SI (**51**) result in regioisomeric ratios of greater than 99:1, but the  $\beta$ -N9-nucleoside tethering Cy<sub>2</sub>SI is isolated in 94% yield (**54**) (Table 2.1) while the  $\beta$ -N9-nucleoside tethering the M<sub>4</sub>SI is isolated in a respectable but

reduced yield of 71% with a slightly higher quantity of  $\alpha$ -nucleosides. Introduction of a methyl group at C3 as in 3-deaz-4-methyladenine (**52**) has a dramatic effect on both the product ratio and isolated yields. The N9:N7 ratio was found to be 6.4:1 for the M<sub>4</sub>SI – protected derivative and substantially better at 20:1 for the Cy<sub>2</sub>SI-protected derivative (**55**). The isolated yield remains better for the Cy<sub>2</sub>SI derivative (85% vs. 59%). The results are promising when compared to chloro and phthalimide at the C6 position of the 3-methyl-3-deazapurine (Table 2.1). Tethering even larger substituents does not substantially alter the results. Using a (4-methoxyphenyl)methanethiol tethered to a 3-methyl-3-deaza-adenine (**53**) would seem to offer a more significant steric challenge for these glycosylation conditions. The N9:N7 ratio is 7:1 and 16:1 – very similar to the results of 3-deaza-4-methyladenine, and the isolated yields follow a similar trend at 57% and 82% (**56**), respectively (Table 2.1). It is also noteworthy that the substrates with Cy<sub>2</sub>SI also produce significantly lower amounts of the undesired  $\alpha$ -isomers.

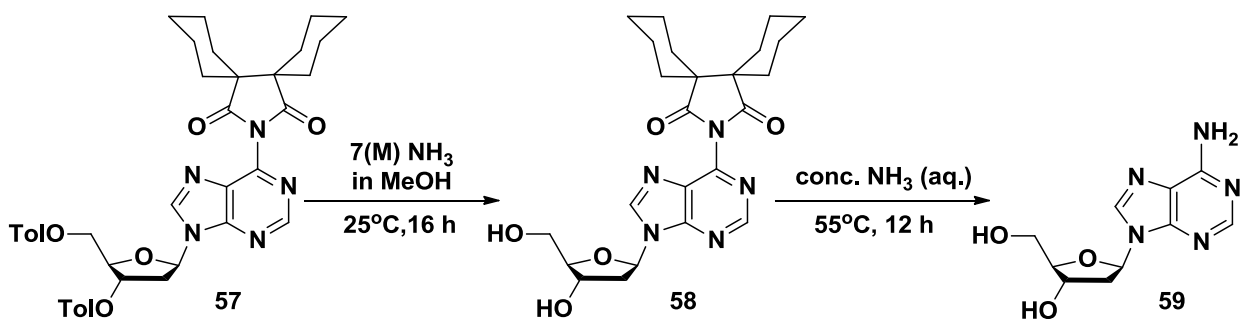
Nucleoside		$\beta$ 9: $\beta$ 7	%yield $\beta$ 9	Total %yield of other isomers
	R = Cl <sup>a</sup>	4.7:1	57	13
	R = Phthalimide <sup>a</sup>	>99:1	36	0
	R = M <sub>4</sub> SI <sup>a</sup>	>99:1	71	6
	R = Cy <sub>2</sub> SI	>99:1	94	3
	R = Cl <sup>a</sup>	1.2:2	39	61
	R = Phthalimide <sup>a</sup>	1:1.1	15	47
	R = M <sub>4</sub> SI <sup>a</sup>	6.4:1	59	27
	R = Cy <sub>2</sub> SI	20:1	85	12
	R = M <sub>4</sub> SI	7:1	57	21
	R = Cy <sub>2</sub> SI	16:1	82	10

**Table 2.1.** Isolated Glycosylation Yields. <sup>a</sup> See Ref. 23.



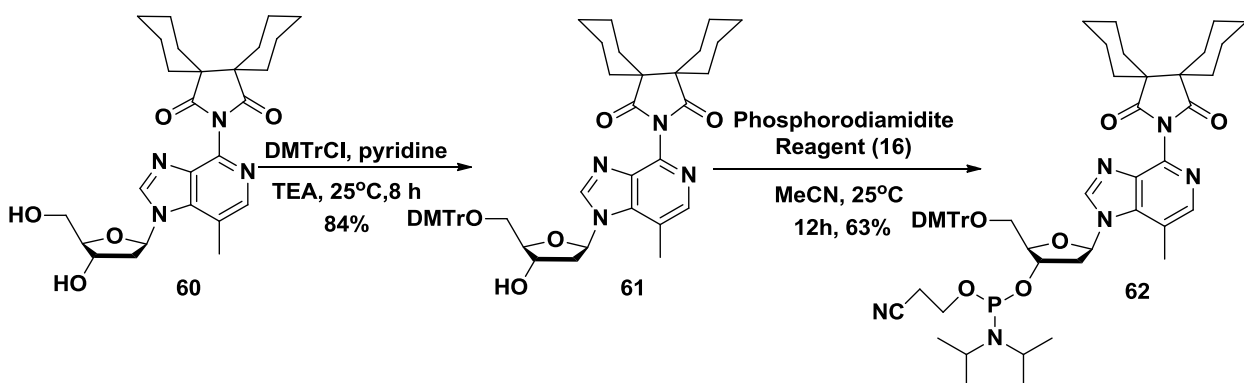
The reason of lower yields of the phthalimide and M<sub>4</sub>SI as protecting groups in the sodium salt methodology may be due to their instability in the basic reaction conditions. These groups fall off in these conditions and as a result the solubility of the corresponding bases decrease which lead to lower overall yields. Removals of these groups also lead to lower regioselectivity toward the desired  $\beta$ -N9 product. On the contrary, the Cy<sub>2</sub>SI is stable in the basic reaction conditions. Hence neither the solubility nor the regioselectivity is compromised.

The hydrolytic stability of this protecting group is also advantageous towards their uses in the automated DNA synthesis. After glycosylation, the toluyl esters of **57** can be selectively removed by 7(M) ammonia in methanol at ambient temperature, while leaving the Cy<sub>2</sub>SI protecting group in place (**58**). Alternatively, under conditions of conc. ammonia (aq.) at 55°C, both the toluyl esters and the diimide protecting group are removed (Scheme 2.18). In the former case the partially protected nucleoside can be further converted to the DMTr-protected phosphoramidite (**62**) suitable for DNA synthetic procedures (Scheme 2.19) with the Cy<sub>2</sub>SI protecting group removed during standard DNA deprotection conditions.



**Scheme 2.18.** Selective removal of toluyl groups and Cy<sub>2</sub>SI group.

Running the reaction in less polar solvents is thought to reduce the number of  $\alpha$ -nucleosides produced during the reaction<sup>24</sup>. Further investigation to develop directing/protecting groups is ongoing to facilitate the glycosylation reactions in non-polar solvents like dichloromethane or dichloroethane.



**Scheme 2.19.** Route to the preparation of Cy<sub>2</sub>SI-protected DMTr-nucleoside phosphoramidite.

In summary, we have developed a new directing/protecting group that provides an effective route to access the 6-aminopurine-2'-deoxynucleosides with high regioselectivity towards the desired  $\beta$ -N9 product over  $\beta$ -N7; especially for the synthesis of valuable C3 substituted purines<sup>25</sup>. X-ray crystallography confirms that portions of both cyclohexyl groups are positioned to sterically block access of the N7 nitrogen so that glycosylation reactions occur with very high regiochemical control at N9. The favorable hydrolytic property of the Cy<sub>2</sub>SI group makes it an ideal protecting group for the phosphoramidite chemistry used in automated DNA synthetic procedures.

## 2.6. Experimental\*

### General Procedures

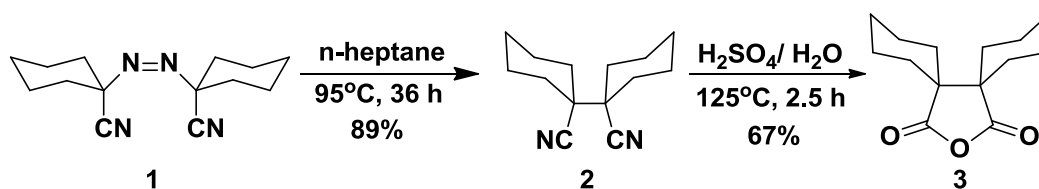
All reactions were carried out under an inert atmosphere with dry solvents under anhydrous conditions unless noted otherwise. Dry tetrahydrofuran (THF), diethyl ether (Et<sub>2</sub>O), *N,N*-dimethylformamide (DMF), pyridine (pyr), acetonitrile (MeCN), and dichloromethane (DCM) were obtained by passing commercially available pre-dried, oxygen-free formulations through activated alumina columns. Dry methanol (MeOH) was obtained by distillation from Mg(OMe)<sub>2</sub>. Yields refer to chromatographically and spectroscopically (<sup>1</sup>H NMR) homogeneous materials, unless otherwise stated. Reagents were purchased at the highest commercial quality and used without further purification, unless otherwise stated. Reactions were monitored by thin-layer chromatography (TLC) carried out on 0.25 mm Silicycle TLG-R10011B-323 60 Å plates using UV light as visualizing agent and ceric ammonium molybdate (CAM) stain or 10% sulphuric acid solution (aq.) and heat as developing agents. Silicycle silica gel (SilicaFlash P60, 60 Å, particle size 40-63 µm) was used for flash column chromatography. Difficult separations were carried out using “chromatospec silica gel” (E. Merck Chromatospec silica gel (60 Å, particle size 15-40 µm). NMR spectra were recorded on Varian VNMRS 400, VNMRS 500, VNMRS 600, or INOVA 500 instruments and calibrated using residual undeuterated solvent (CDCl<sub>3</sub>: δ<sub>H</sub> = 7.26 ppm, δ<sub>C</sub> = 77.16 ppm, methanol-

---

\* The compound numbers are reset here for the rest of the chapter.

*d*<sub>4</sub>:  $\delta\text{H} = 3.31$  ppm,  $\delta\text{C} = 49.00$  ppm) as an internal reference. The following abbreviations are used to designate the multiplicities: s = singlet, d = doublet, t = triplet, q = quartet, quin = quintet, m = multiplet, br = broad. High-resolution mass spectra (HRMS) were recorded on a Waters LCT Classic or JEOL AccuTOF mass spectrometer using ESI (electrospray ionization) or DART (direct analysis in real time).

### I. Synthesis of Bicyclohexyl-1,1'-dicarboxylic anhydride



**Scheme 2.20.** Synthesis of Bicyclohexyl-1,1'-dicarboxylic acid cyclic-anhydride (Cy<sub>2</sub>SA)

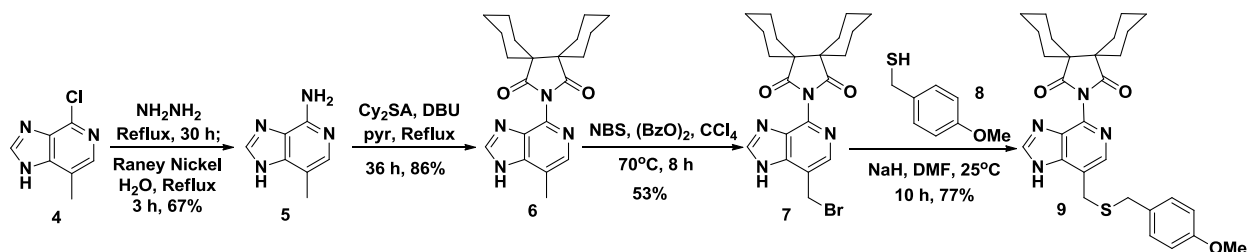
**[1,1'-bi(cyclohexane)]-1,1'-dicarbonitrile (2).** To a flask of containing **1** (5 g, 20.463 mmol) was added n-heptane (28 mL). The mixture was heated to 95°C for 36 h. The solution was allowed to cool to room temperature. The suspension was filtered and the residue was washed with hexanes (3×75 mL) to obtain light yellowish white product (4.129 g, 89%). **2**: <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>)  $\delta$  2.12 (d, *J* = 11.0 Hz, 4H), 1.90 - 1.73 (m, 6H), 1.61 (d, *J* = 13.3 Hz, 8H), 1.25 - 1.16 (m, 2H); <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>)  $\delta$  120.42, 46.45, 31.01, 24.87, 23.16; HRMS (ESI-TOF) calcd for [M + H]<sup>+</sup> 234.19702, found: 234.19744.

**Bicyclohexyl-1,1'-dicarboxylic acid cyclic-anhydride (3).** A solution of H<sub>2</sub>SO<sub>4</sub> (44 mL) in water (19 mL) was prepared and cooled to room temperature. To a flask

containing **2** (7.879 g, 36.425 mmol) was added the aq. H<sub>2</sub>SO<sub>4</sub> solution. It was steadily heated at 115°C for 1h and 15min. Then the temperature was raised to 125°C and maintained for 1h and 15min. The solution was allowed to cool to room temperature. The resulting dark solution was extracted with diethyl ether (3×200 mL), the combined organic extracts were washed water (3×100 mL) followed by saturated NaHCO<sub>3</sub> (3×100 mL), dried over anhydrous MgSO<sub>4</sub>, filtered and concentrated *in vacuo*. The resulting slightly yellowish residue was purified by flash column chromatography (1:30 ethyl acetate: hexanes) to yield white amorphous powder (2.883 g, 67%). **3**: *R*<sub>f</sub>= 0.42 (silica gel, 1:20 ethyl acetate:hexane); <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 1.86 - 1.60 (m, 14H), 1.44 - 1.33 (m, 4H), 1.17 - 1.05 (m, 2H); <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 174.52, 52.95, 29.12, 25.39, 22.08; HRMS (DART-TOF) calcd for [M + H]<sup>+</sup> 237.14907, found: 237.14909.

## II. Synthesis of 3-deaza-3-substituted Purine Substrates

The 3-deaza-3-substituted substrates were prepared from the 3-deaza-3-methyl analog **5**.



**Scheme 2.21.** Synthesis 3-deaza-3-substituted substrates

**3-deaza-3-methyl-6-amino protected purine (6).** To a flask containing **5** (1.2 g, 8.099 mmol) and Cy<sub>2</sub>SA (5.742 g, 24.297 mmol) was added pyridine (80 mL) and DBU (3.63

mL, 24.297 mmol). The mixture was heated to reflux for 36 h. Volatiles were removed *in vacuo* and the residue co-evaporated with toluene (3 x 10 mL) to remove residual pyridine. The resulting brown oil was purified by flash column chromatography eluting with 1:25 MeOH:DCM to obtain of the compound **6** (2.552 g, 86%) as a white amorphous powder. **6**:  $R_f$  = 0.32 (silica gel, 1:20 MeOH:DCM);  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.00 (s, 1H), 7.36 (s, 1H), 2.21 (s, 3H), 1.98 (dd,  $J$  = 25.2, 11.7 Hz, 8H), 1.79 - 1.69 (m, 6H), 1.60 - 1.53 (m, 4H), 1.25 - 1.17 (m, 2H);  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  181.46, 142.94, 140.96, 139.58, 135.77, 135.69, 119.22, 52.07, 30.21, 25.65, 22.47, 14.01; HRMS (DART-TOF) calcd for  $[\text{M} + \text{H}]^+$  367.21340, found: 367.21464..

**3-deaza-3-bromomethyl-6-amino protected purine (7)**. To a flask containing **6** (1.5 g, 4.093 mmol) was added *N*-bromosuccinimide (1.092 g, 6.140 mmol, recrystallized from hot  $\text{H}_2\text{O}$ ), benzoyl peroxide (99 mg, 0.409 mmol), and 120 mL of  $\text{CCl}_4$  (reagent grade, not anhydrous). The mixture was heated to reflux with stirring. After 5 hr additional NBS (364 mg, 2.047 mmol) and  $(\text{BzO})_2$  (99 mg, 0.409 mmol) were added. After 8 hr (total), all the starting material was consumed. The reaction was allowed to cool to rt and filtered to removed succinimide. Volatiles were removed *in vacuo* and the residue purified by flash column chromatography eluting with 3:2 ethyl acetate: hexane to afford the desired product as a white foam (0.966 g, 53% ). **7**:  $R_f$  = 0.23 (silica gel, 1:1 ethyl acetate: hexane);  $^1\text{H}$  NMR (500 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  8.44 (s, 1H), 8.41 (s, 1H), 4.98 (s, 2H), 2.09 – 1.93 (m, 8H), 1.75 – 1.70 (m, 6H), 1.63 (dd,  $J$  = 13.0, 9.2 Hz, 4H), 1.32 - 1.27 (m, 3.5 Hz, 2H);  $^{13}\text{C}$  NMR (126 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  181.59, 146.46, 141.70, 139.16, 122.82, 106.42, 53.17, 31.01, 26.67,

26.08, 23.55; HRMS (DART-TOF) calcd for  $[M+H]^+$  445.12391, found: 445.12421.

**3-deaza-3-(4-methoxyphenyl)methanethiolmethyl-6-amino protected purine (9).**

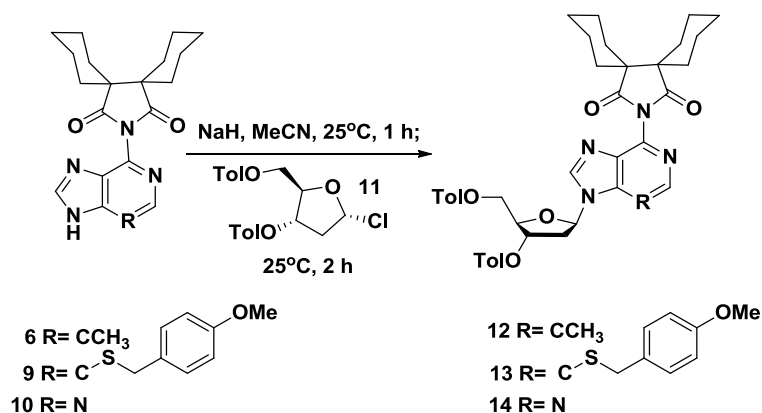
To a flask containing sodium hydride (53 mg, 2.19 mmol) was added DMF (15 mL) followed by the thiol reagent **8** (0.31 mL, 2.19 mmol). The mixture was stirred at 25°C for 30 min. The compound **7** (0.65 g, 1.460 mmol) was added and allowed to stir for another 10 h at room temperature. The TLC indicated the reaction was complete. Volatiles were removed *in vacuo* and the residue purified by flash column chromatography eluting with 1:40 MeOH:DCM to afford the desired product as a white foam (0.583 g, 77%). **9**:  $R_f$  = 0.35 (silica gel, 1:25 MeOH:DCM);  $^1\text{H}$  NMR (500 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  8.33 (s, 1H), 8.14 (s, 1H), 7.16 (d,  $J$  = 8.5 Hz, 2H), 6.87 – 6.80 (m, 2H), 4.01 (s, 2H), 3.78 (s, 3H), 3.64 (d,  $J$  = 8.3 Hz, 2H), 2.11 – 1.94 (m, 8H), 1.79 – 1.70 (m, 6H), 1.63 (td,  $J$  = 13.0, 3.9 Hz, 4H), 1.33 – 1.28 (m, 2H);  $^{13}\text{C}$  NMR (126 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  180.39, 158.81, 144.46, 139.86, 131.28, 129.66, 129.40, 113.94, 113.47, 54.25, 51.64, 34.69, 29.61, 29.14, 25.27, 22.14; HRMS (DART-TOF) calcd for  $[M+H]^+$  519.2430, found: 519.2446.

**6-Amino protected adenine (10).** To a flask containing adenine (0.25 g, 1.85 mmol) and  $\text{Cy}_2\text{SA}$  (1.312 g, 5.55 mmol) was added pyridine (20 mL) and DBU (0.83 mL, 5.55 mmol). The mixture was heated to reflux for 20 h. Volatiles were removed *in vacuo* and the residue co-evaporated with toluene (3×10 mL) to remove residual pyridine. The resulting brown oil was purified by flash column chromatography eluting with 1:30 MeOH:DCM to obtain of the compound **10** as a white amorphous powder (0.582 g, 89%). **10**:  $R_f$  = 0.42 (silica gel, 1:20 MeOH:DCM);  $^1\text{H}$  NMR (500 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  8.93

(s, 1H), 8.53 (s, 1H), 2.00 - 1.90 (m, 8H), 1.77 - 1.67 (m, 6H), 1.59 (td, J = 13.5, 4.1 Hz, 4H), 1.30 - 1.20 (m, 2H);  $^{13}\text{C}$  NMR (126 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  180.63, 153.22, 147.68, 144.61, 53.50, 30.87, 26.62, 23.50; HRMS (DART-TOF) calcd for  $[\text{M} + \text{H}]^+$  354.19300, found: 354.19345.

### III. Synthesis of the nucleosides of 6, 9 and 10

The nucleosides of the protected nucleobases 6, 9 and 10 were prepared as shown below.



**Scheme 2.22.** Glycosylation reactions of the protected nucleobases

**General Procedure:** To a flask containing protected nucleobase and NaH (1.5 eqvt.), DMF (0.05 (M) solution) was added and stirred for 30 min at room temperature. The sugar (**11**) (1.8 eqvt.) was added to the mixture and stirred for another 2 hr at room temperature. The resulting cloudy solution was filtered through celite and rinsed with acetone. The filtrate was concentrated in vacuo and purified by column chromatography to yield the white foam.



Nucleoside	$\beta 9:\beta 7$	%yield $\beta 9$	Total %yield of other isomers
<b>12</b>	<b>20:1</b>	<b>85</b>	<b>12</b>
<b>13</b>	<b>16:1</b>	<b>82</b>	<b>10</b>
<b>14</b>	<b>&gt;99:1</b>	<b>94</b>	<b>3</b>

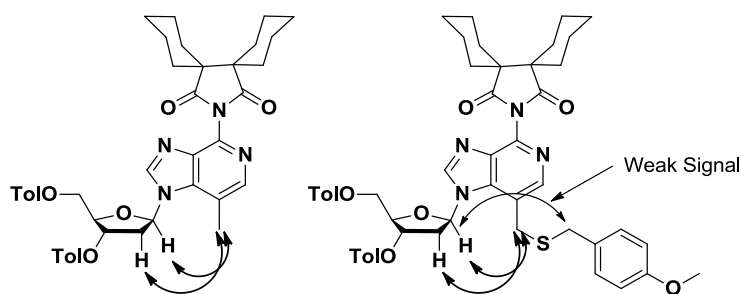
Characterization of **12**:  $R_f$  = 2.20 (silica gel, 1:3 ethyl acetate:hexanes). Purified by gradient column chromatography ranging from 1:4 to 2:3 ethyl acetate:hexanes;  $^1\text{H}$  NMR (500 MHz,  $\text{cdcl}_3$ )  $\delta$  8.24 (s, 1H), 8.22 (s, 1H), 7.95 – 7.92 (m, 2H), 7.87 – 7.84 (m, 2H), 7.30 – 7.27 (m, 2H), 7.24 – 7.21 (m, 2H), 6.63 (t,  $J$  = 6.8 Hz, 1H), 5.71 – 5.67 (m, 1H), 4.68 – 4.57 (m, 3H), 2.80 (qd,  $J$  = 7.9, 4.3 Hz, 2H), 2.70 (d,  $J$  = 0.7 Hz, 3H), 2.43 (d,  $J$  = 5.2 Hz, 3H), 2.40 (s, 3H), 2.08 – 1.94 (m, 8H), 1.77 – 1.66 (m, 6H), 1.52 (qd,  $J$  = 13.1, 3.9 Hz, 4H), 1.18 (tdd,  $J$  = 12.5, 8.1, 4.4 Hz, 2H).;  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  180.31, 166.26, 165.93, 144.80, 144.33, 143.50, 141.74, 139.46, 137.95, 137.21, 129.88, 129.76, 129.48, 126.65, 126.36, 117.80, 85.41, 82.82, 74.62, 63.96, 52.13, 39.54, 30.41, 29.99, 25.90, 22.61, 21.87, 21.78, 15.83; HRMS (ESI-TOF) calcd for  $[\text{M} + \text{H}]^+$  741.3264, found: 741.3242.

Characterization of **13**:  $R_f$  = 2.26 (silica gel, 1:3 ethyl acetate:hexanes). Purified by gradient column chromatography ranging from 1:5 to 3.2:10 ethyl acetate:hexanes;  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.20 (d,  $J$  = 6.2 Hz, 1H), 8.16 (s, 1H), 7.99 - 7.95 (m, 2H), 7.86 - 7.83 (m, 2H), 7.28 - 7.24 (m, 2H), 7.24 - 7.20 (m, 4H), 6.85 - 6.82 (m, 2H), 6.74 (t,  $J$  = 6.6 Hz, 1H), 5.69 - 5.65 (m, 1H), 4.61 - 4.53 (m, 2H), 4.50 (dd,  $J$  = 7.2, 3.9 Hz, 1H), 4.18 - 4.14 (m, 1H), 3.78 (s, 3H), 3.64 (d,  $J$  = 1.8 Hz, 2H), 2.82 - 2.76 (m, 2H), 2.44 (s, 3H), 2.40 (s, 3H), 2.03 (dd,  $J$  = 18.2, 8.0 Hz, 8H), 1.73 (dd,  $J$  = 32.1, 12.4 Hz, 6H), 1.56 – 1.48 (m, 4H), 1.22 - 1.17 (m, 2H);  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  180.25,

166.26, 165.97, 159.00, 144.75, 144.30, 143.52, 142.02, 139.32, 139.16, 137.92, 130.27, 130.01, 129.78, 129.48, 129.27, 126.73, 126.43, 118.77, 114.22, 85.68, 82.52, 74.29, 63.91, 55.42, 52.24, 38.68, 35.43, 30.50, 30.02, 29.94, 25.91, 22.64, 21.89, 21.82; HRMS (ESI-TOF) calcd for  $[M+Na]^+$  893.3560, found: 893.3564.

Characterization of **14**:  $R_f$  = 2.24 (silica gel, 1:3 ethyl acetate:hexanes). Purified by gradient column chromatography ranging from 1:4 to 3.5:10 ethyl acetate:hexanes;  $^1H$  NMR (500 MHz,  $cdCl_3$ )  $\delta$  8.93 (s, 1H), 8.29 (s, 1H), 8.00 – 7.95 (m, 2H), 7.92 – 7.88 (m, 2H), 7.28 (d,  $J$  = 8.0 Hz, 2H), 7.24 (t,  $J$  = 7.5 Hz, 2H), 6.60 (dd,  $J$  = 8.4, 5.7 Hz, 1H), 5.81 (d,  $J$  = 6.2 Hz, 1H), 4.73 (td,  $J$  = 5.9, 2.9 Hz, 1H), 4.66 (dt,  $J$  = 7.4, 4.4 Hz, 2H), 3.15 (ddd,  $J$  = 14.5, 8.5, 6.3 Hz, 1H), 2.85 (ddd,  $J$  = 14.2, 5.7, 1.9 Hz, 1H), 2.44 (s, 3H), 2.40 (s, 3H), 2.04 – 1.95 (m, 8H), 1.81 – 1.69 (m, 6H), 1.59 – 1.49 (m, 4H), 1.19 (dtd,  $J$  = 12.7, 9.0, 3.7 Hz, 2H).  $^{13}C$  NMR (126 MHz,  $CDCl_3$ )  $\delta$  179.10, 166.28, 166.00, 153.32, 152.67, 145.45, 144.66, 144.25, 144.02, 130.34, 129.94, 129.79, 129.41, 126.79, 126.49, 85.20, 83.30, 75.13, 64.10, 53.94, 52.63, 37.96, 30.07, 29.94, 29.40, 25.78, 22.51, 21.86, 21.78; HRMS (ESI-TOF) calcd for  $[M+Na]^+$  728.3060, found: 728.3028.

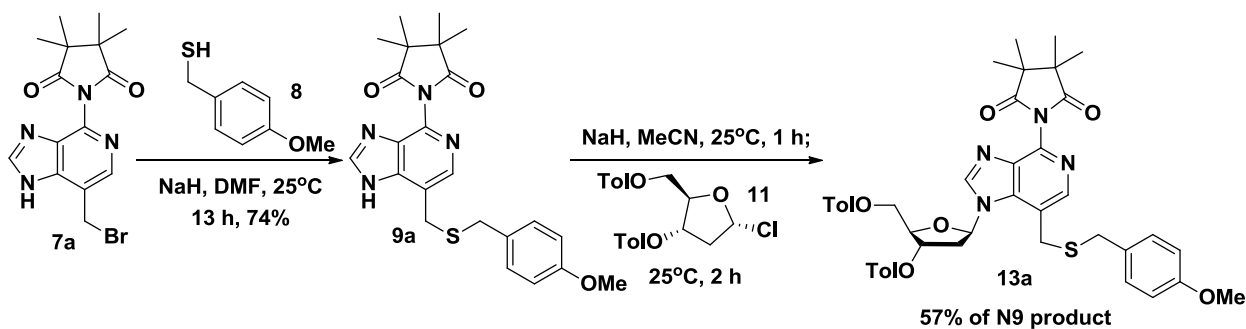
Determination of regioselectivity of **12** and **13**: Regioselectivity was determined by NOESY experiment. The important interactions (Figure 1) has been indicated. There were no interactions of any of the hydrogens of the  $Cy_2SI$  with the  $C1'$  hydrogen. This proves that the **12** and **13** were the N9 products, and not N7.



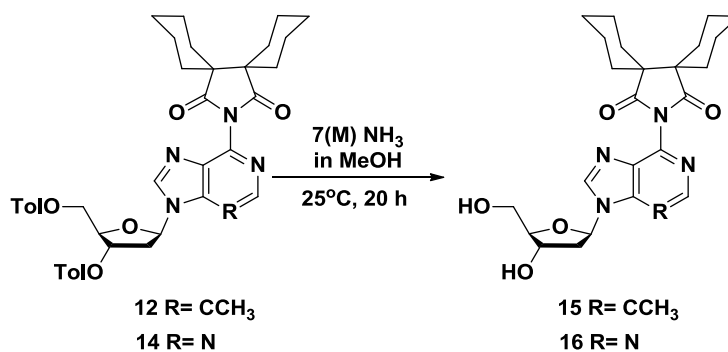
**Figure 1:** NOESY signals from NOESY experiment for compound **12** and **13**.

### IIIa. Details of M<sub>4</sub>SI (tetramethylsuccinimide) derivatives of **7**, **9** and **13**

**9a** was synthesized from **7a** following the similar procedure used to convert **7** to **9**. The reaction time and yield were 13 h and 74% respectively. General procedure for glycosylation was performed to couple **9a** with **11** to yield 57% of the desired  $\beta$ -N9 product.



### IV. Selective removal of toluyl groups



**Scheme 2.23.** Removal of toluyl groups without cleaving the Cy<sub>2</sub>SI group

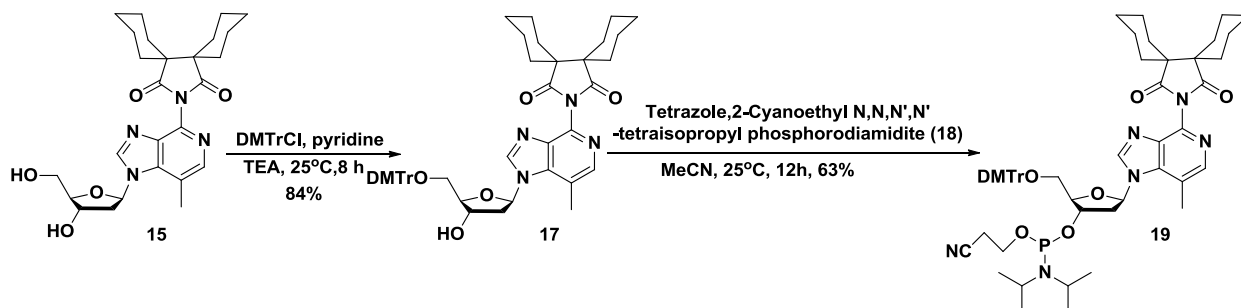
**General Procedure:** To a flask containing the nucleoside was added 7(M) NH<sub>3</sub> containing methanol (0.05(M) solution). It was stirred at room temperature for 20 hr. Solvent was removed *in vacuo* and resulting crude was purified by flash column chromatography to yield white solid.

Yield of 15 and 16 was found to be 89% and 92% respectively.

Characterization of **15**: <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD) δ 8.68 (s, 1H), 8.16 (s, 1H), 6.71 (t, J = 6.3 Hz, 1H), 4.58 (dt, J = 6.1, 4.3 Hz, 1H), 4.06 (q, J = 3.6 Hz, 1H), 3.73 (ddd, J = 33.6, 12.1, 3.7 Hz, 2H), 2.83 – 2.75 (m, 1H), 2.798 (s, 3H), 2.58 (ddd, J = 13.4, 6.0, 4.5 Hz, 1H), 2.16 – 1.92 (m, 8H), 1.72 (dd, J = 20.7, 9.2 Hz, 6H), 1.61 (t, J = 12.2 Hz, 4H), 1.33 – 1.24 (m, 2H); <sup>13</sup>C NMR (101 MHz, CD<sub>3</sub>OD) δ 180.46, 143.89, 141.74, 139.81, 136.61, 119.61, 87.93, 85.46, 70.59, 61.18, 51.65, 40.44, 29.54, 25.25, 22.13, 14.26; HRMS (DART-TOF) calcd for [M+ H]<sup>+</sup> 483.26074, found: 483.25878.

Characterization of **16**: <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD) δ 8.93(s, 1H), 8.75 (s, 1H), 6.59 (t, J = 6.6 Hz, 1H), 4.64 - 4.60 (m, 1H), 4.06 (q, J = 3.6 Hz, 1H), 3.83 (dd, J = 12.1, 3.5 Hz, 1H), 3.76 (dd, J = 12.1, 4.1 Hz, 1H), 2.91 - 2.85 (m, 1H), 2.52 (ddd, J = 13.6, 6.4, 3.8 Hz, 1H), 2.01 - 1.91 (m, 8H), 1.72 (dd, J = 21.1, 9.7 Hz, 6H), 1.58 (dd, J = 20.3, 7.6 Hz, 4H), 1.30 - 1.22 (m, 2H); <sup>13</sup>C NMR (101 MHz, CD<sub>3</sub>OD) δ 179.19, 153.20, 151.69, 144.58, 130.36, 88.20, 85.21, 71.05, 61.69, 52.25, 39.90, 29.37, 25.20, 22.07; HRMS (DART-TOF) calcd for [M+ H]<sup>+</sup> 470.24034, found: 470.23837.

## V. Synthesis of phosphoramidite from 15



**Scheme 2.24.** Synthesis of DMTr-protected phosphoramidite of 15

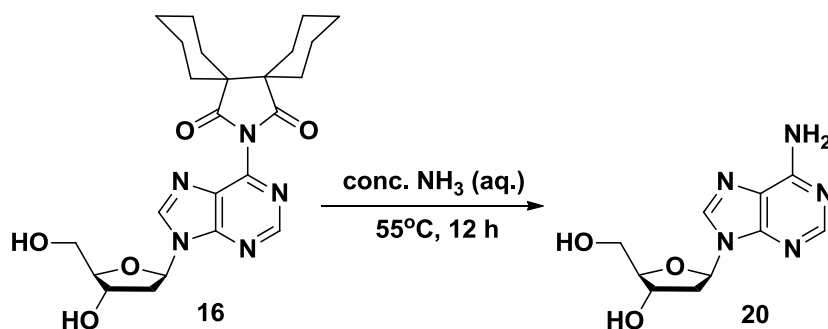
**Compound 17:** To a flask containing **15** (35 mg, 0.073 mmol) dissolved in pyridine (0.8 mL) was added TEA (14  $\mu$ L, 0.102 mmol), DMTrCl (30 mg, 0.087 mmol), and DMAP (0.9 mg, 0.007 mmol). The reaction was allowed to stir at rt for 5 h. TLC analysis indicated that some starting material remained, and so additional DMTrCl (5 mg, 0.015 mmol) was added. TLC after another 8 h indicated complete consumption of starting material. Volatiles were removed *in vacuo* and the solid residue was then purified by flash chromatography (silica gel was washed with 1% TEA in 1:50 MeOH:DCM) eluting with 1% TEA in 1:50 MeOH:DCM to afford **22** as a white foam (48 mg, 84%). **17**:  $R_f$  = 0.23 (silica gel, 1:50 MeOH:DCM);  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.16 (s, 1H), 8.11 (s, 1H), 7.41 - 7.35 (m, 2H), 7.30 - 7.18 (m, 7H), 6.81 (dd,  $J$  = 8.9, 1.6 Hz, 4H), 6.47 (t,  $J$  = 6.3 Hz, 1H), 4.51 (dd,  $J$  = 9.5, 4.3 Hz, 1H), 4.08 (dd,  $J$  = 8.7, 4.4 Hz, 1H), 3.78 (s, 6H), 3.38 (dd,  $J$  = 10.2, 4.4 Hz, 1H), 3.28 (dd,  $J$  = 10.2, 5.1 Hz, 1H), 2.63 (s, 3H), 2.49 (tdd,  $J$  = 13.1, 11.7, 6.0 Hz, 2H), 2.13 - 1.93 (m, 8H), 1.77 - 1.63 (m, 6H), 1.56 - 1.46 (m, 4H), 1.38 (dd,  $J$  = 13.9, 6.8 Hz, 2H);  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  180.31, 158.65, 144.42, 143.05, 142.00, 139.33, 137.61, 137.18, 135.52, 135.42, 130.03, 129.98, 128.03, 127.98, 127.01, 117.98, 113.29, 86.75, 85.74, 84.90, 72.18, 63.67, 55.24, 51.99, 41.35, 30.27, 29.93, 25.78, 22.50, 15.73; HRMS (ESI-TOF)

calcd for  $[M+H]^+$  785.3914, found: 785.3941.

**Compound 19:** To a flask containing **17** (35 mg, 0.045 mmol) dissolved in MeCN (0.9 mL) was added tetrazole (3 mg, 0.045 mmol) followed by 2-cyanoethyl N,N,N',N'-tetraisopropylphosphane(**18**) (30  $\mu$ L, 0.089 mmol) dropwise. The mixture was allowed to stir at rt for 12 h, after which TLC indicated that the reaction was complete. Water (0.1 mL) was added and stirring continued for 10 min. The reaction mixture was diluted with DCM (5 mL) and washed with aq. saturated  $\text{NaHCO}_3$  solution (2 $\times$ 2 mL) and brine (1 $\times$ 2 mL). The organic layer was separated and dried with  $\text{Na}_2\text{SO}_4$ , then filtered and dried *in vacuo*. It was dissolved in 0.2 mL of DCM. To this 2 mL of hexane was added when the product crushed out as white solid. The liquid phase was removed. This process of dissolving and precipitation was carried out 5 times. Then the white residue was purified by short flash column chromatography (silica gel washed with 1% TEA in 0.5:9.5 MeOH:DCM) eluting with 0.5:100 MeOH:DCM to afford **19** as a white foam (28 mg, 63%, 1:1 mixture of diastereomers). **19**:  $R_f$  = 0.29 (silica gel, 1:50 MeOH:DCM);  $^{31}\text{P}$  NMR (202 MHz,  $\text{CDCl}_3$ )  $\delta$  149.48, 149.07.

HRMS (ESI-TOF) calcd for  $[M+H]^+$  985.4993, found: 985.4982.

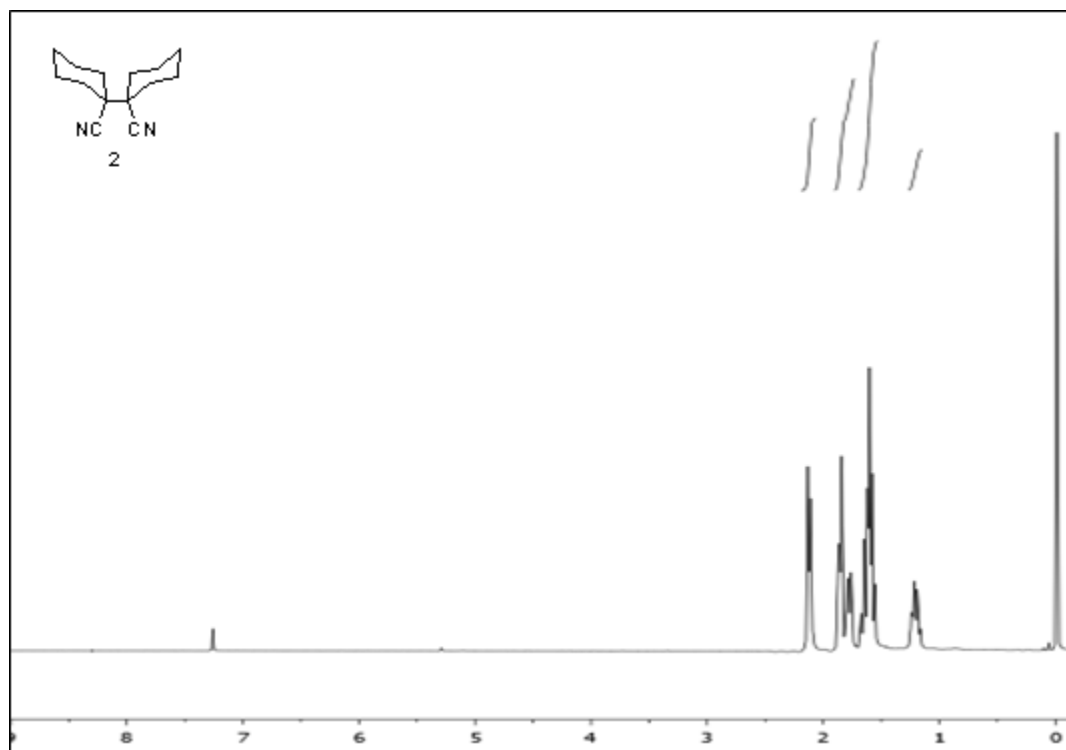
## VI. Removal of the directing/ protecting group



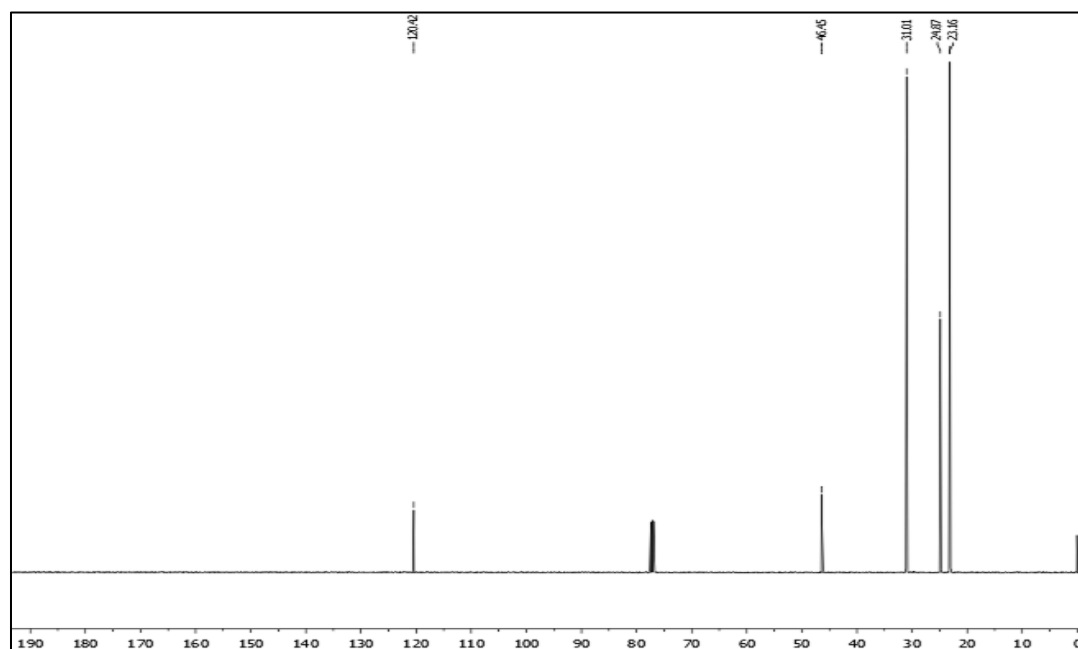
**Compound 20:** The nucleoside **16** (20 mg, 0.043 mol) was taken in a vial. To this

conc.  $\text{NH}_3$  (aq.) (2 mL) was added. It was allowed to remain in the water bath at  $55^\circ\text{C}$  for 12 h. The water was removed *in vacuo* and the obtained crude was purified by HPLC to obtain compound 20. **20**:  $^1\text{H}$  NMR (500 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  8.34 (s, 1H), 8.20 (s, 1H), 6.45 (dd,  $J = 7.9, 6.1$  Hz, 1H), 4.60 (dt,  $J = 5.7, 2.6$  Hz, 1H), 4.09 (dd,  $J = 5.7, 2.9$  Hz, 1H), 3.82 (ddd,  $J = 50.8, 12.3, 3.1$  Hz, 2H), 2.83 (ddd,  $J = 13.6, 7.9, 5.8$  Hz, 1H), 2.43 (ddd,  $J = 13.4, 6.0, 2.7$  Hz, 1H);  $^{13}\text{C}$  NMR (126 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  157.52, 153.49, 149.95, 141.54, 120.86, 89.91, 87.15, 73.06, 63.65, 41.57; HRMS (DART-TOF) calcd for  $[\text{M} + \text{H}]^+$  252.10966, found: 252.11006.

$^1\text{H}$  NMR of Compound **2**

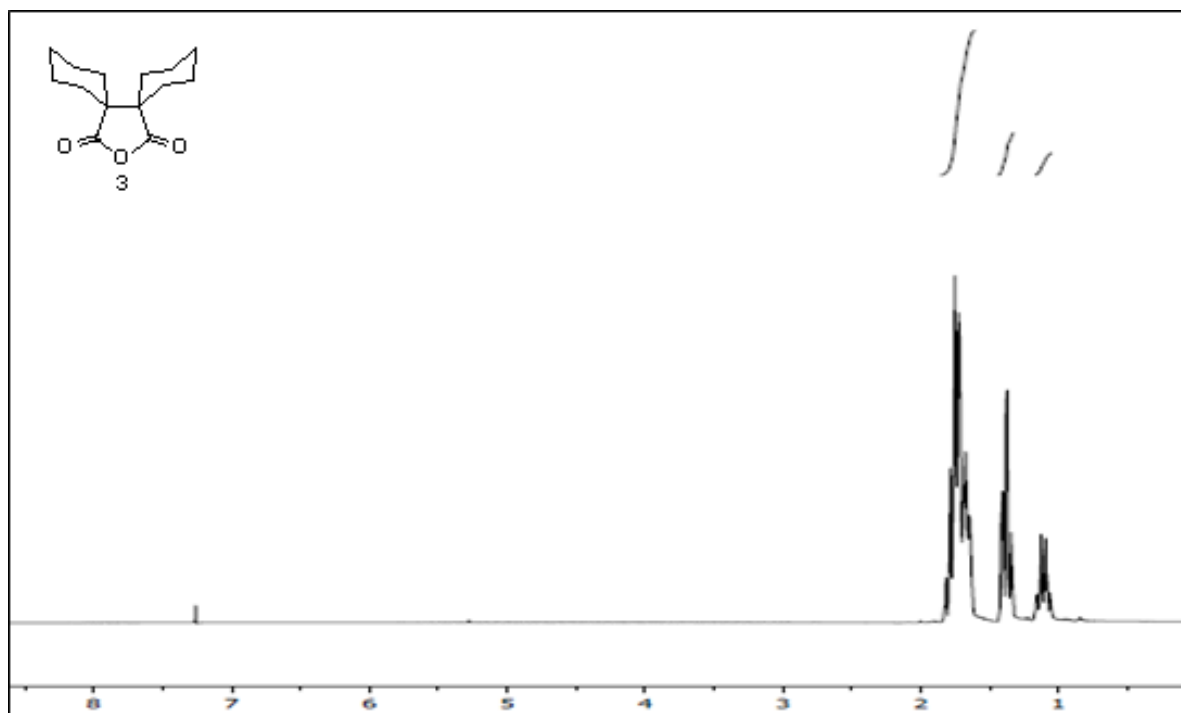


$^{13}\text{C}$  NMR spectrum of Compound **2**

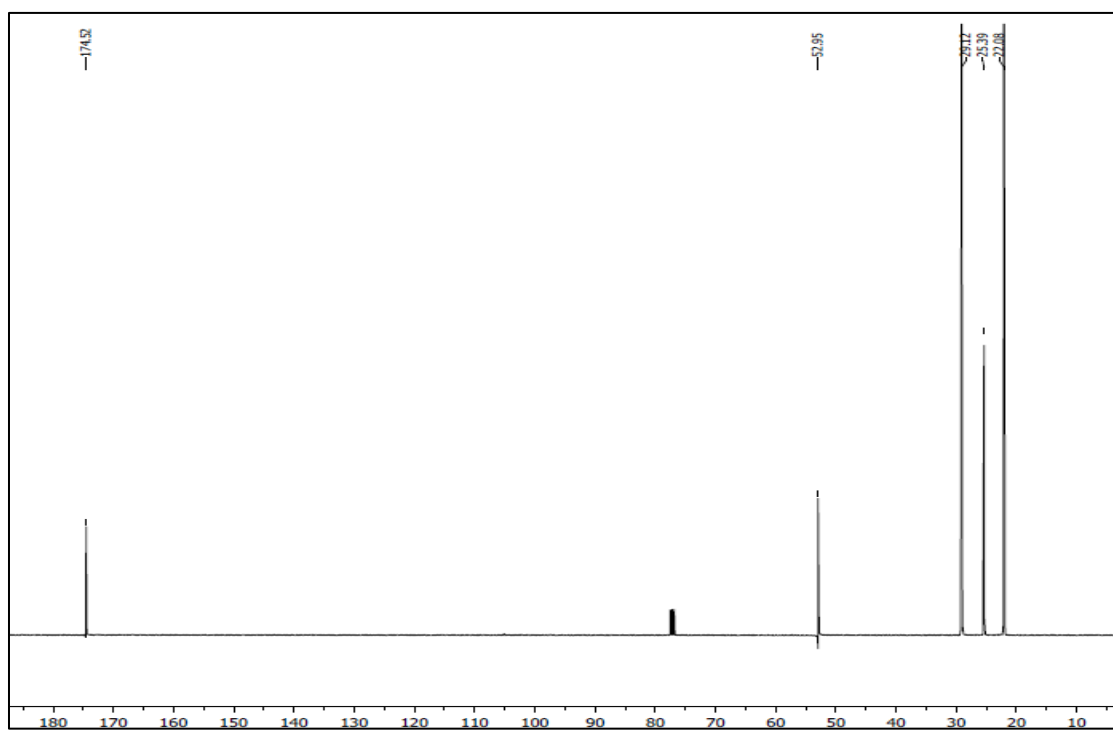




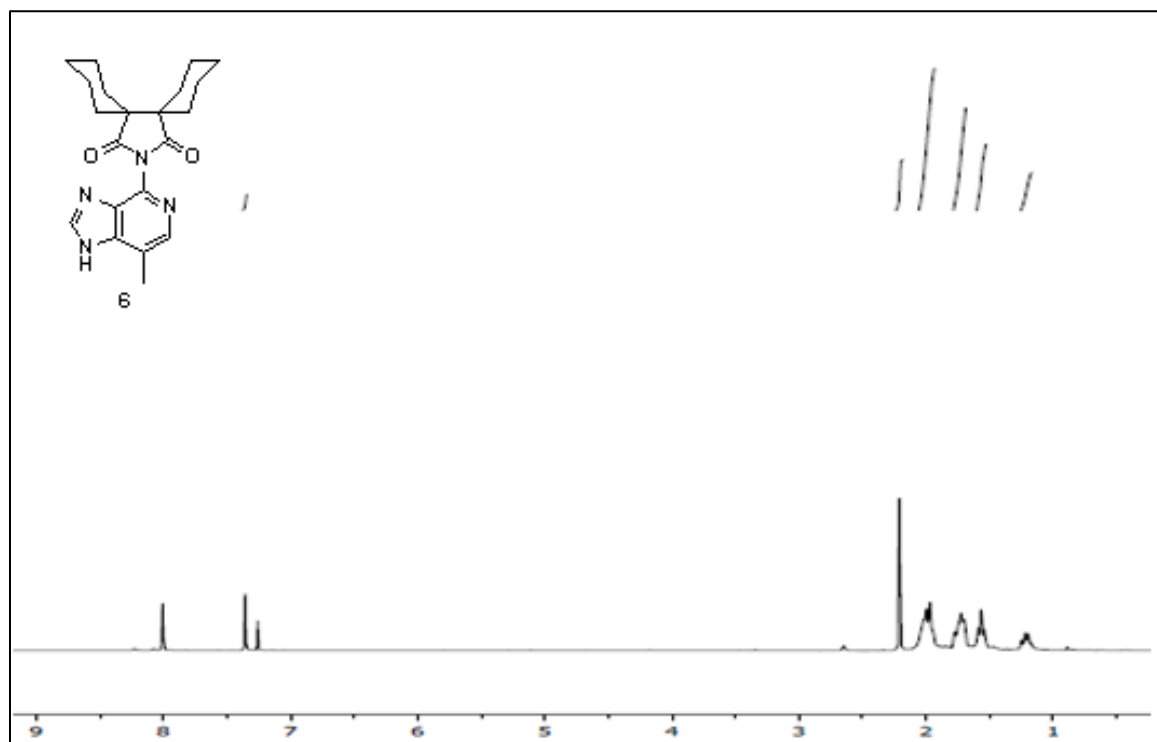
$^1\text{H}$  NMR of Compound **3**



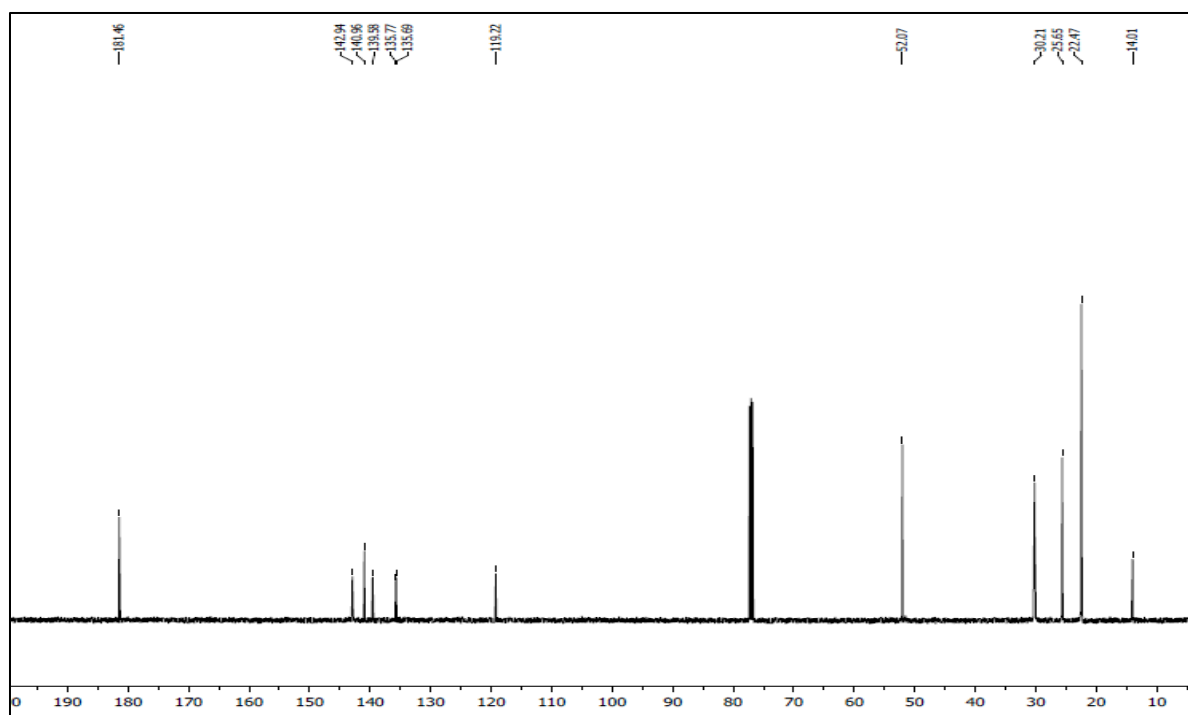
$^{13}\text{C}$  NMR of Compound **3**



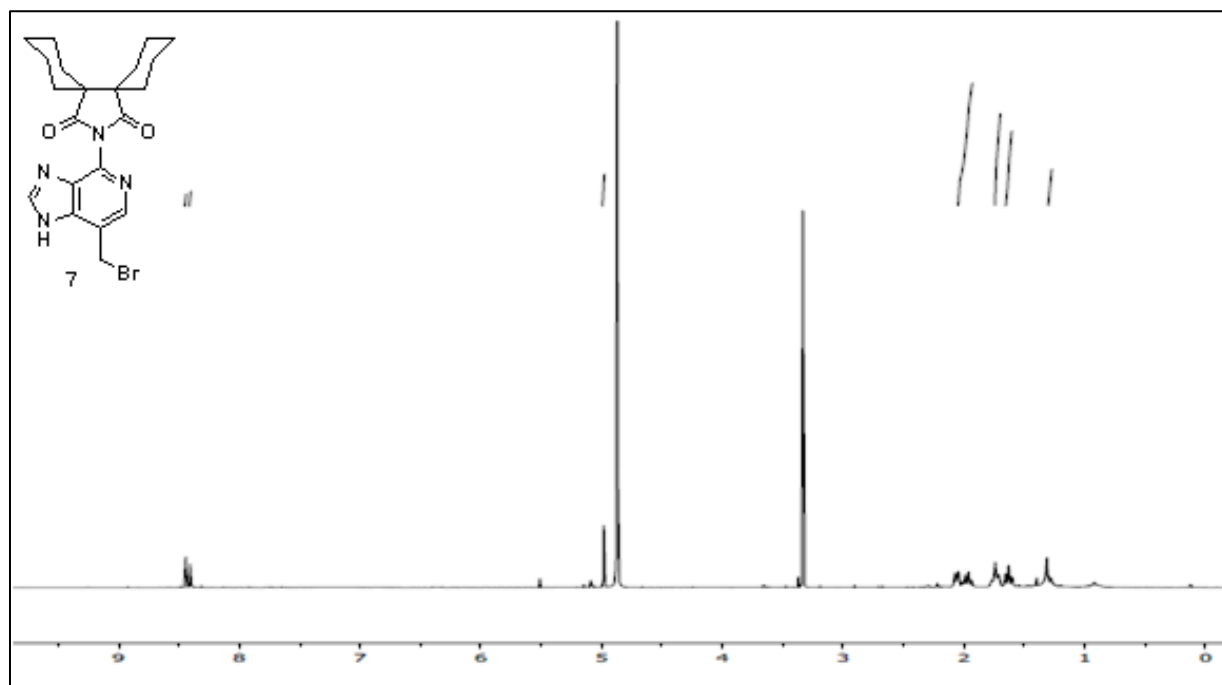
$^1\text{H}$  NMR of Compound **6**



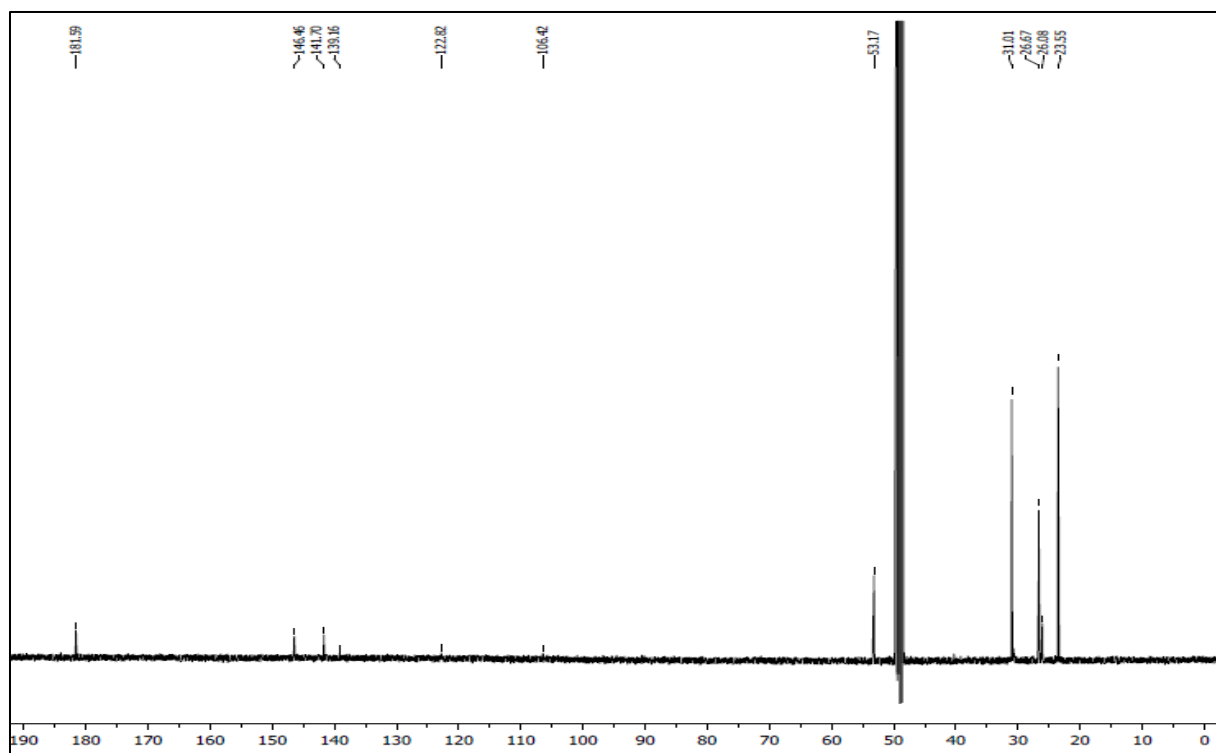
$^{13}\text{C}$  NMR of Compound **6**



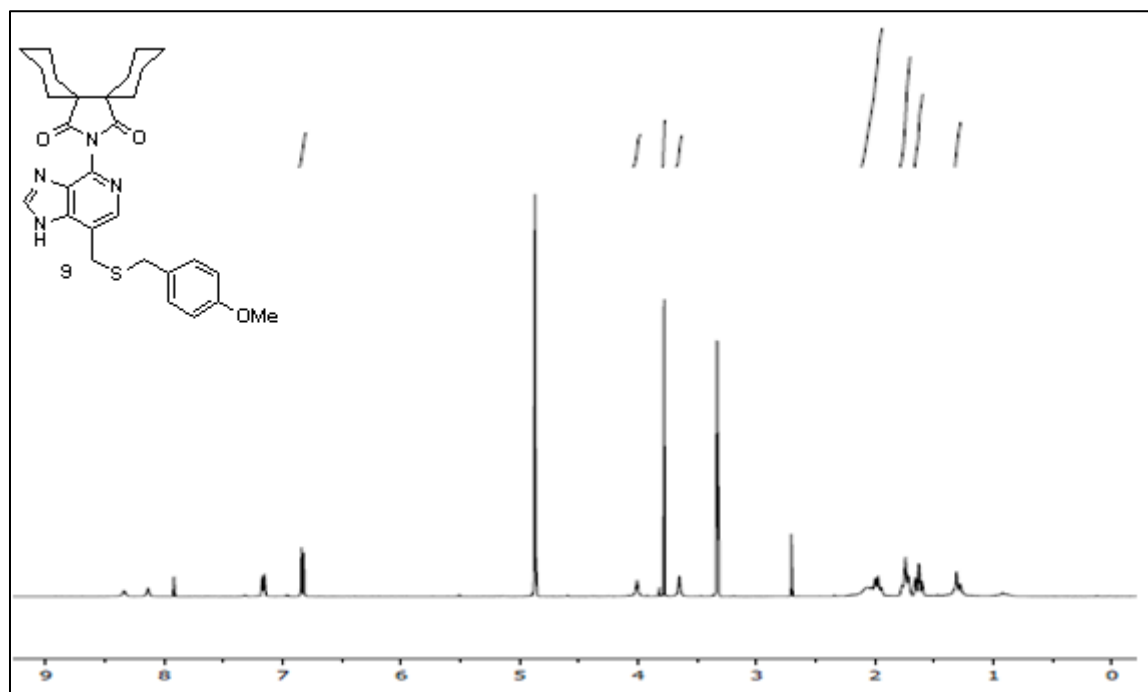
$^1\text{H}$  NMR of Compound 7



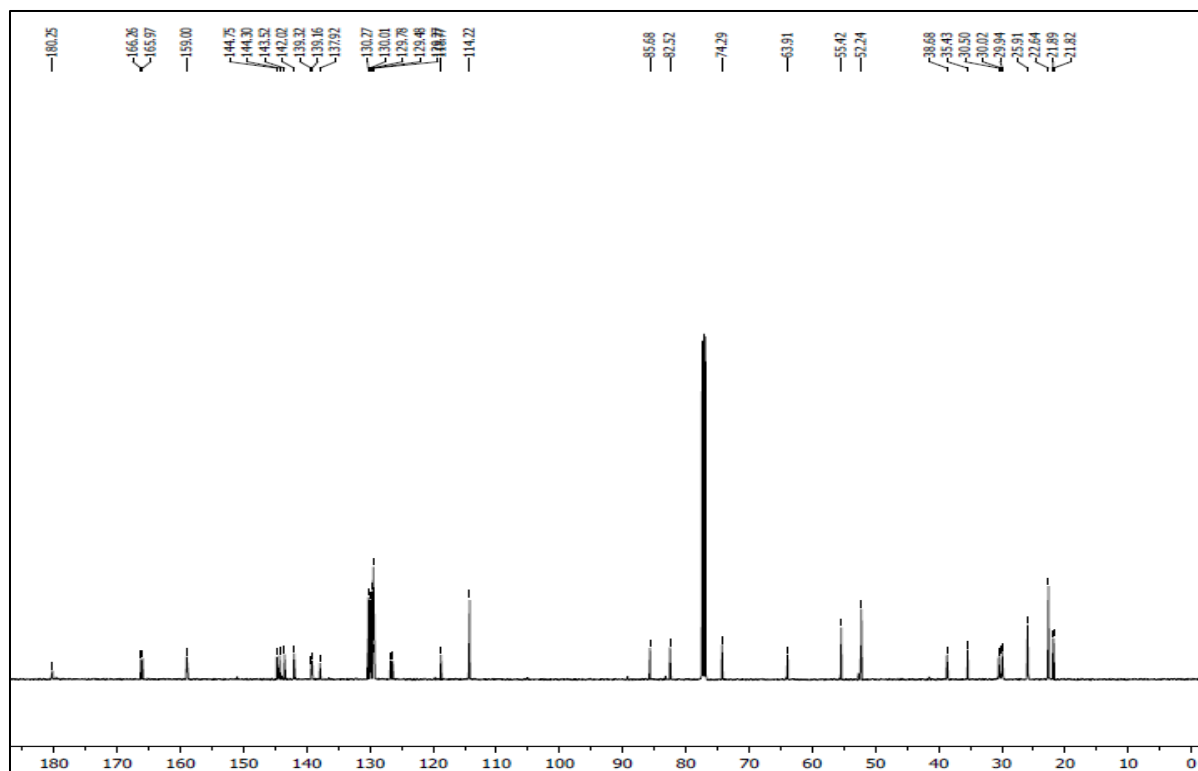
$^{13}\text{C}$  NMR of Compound 7



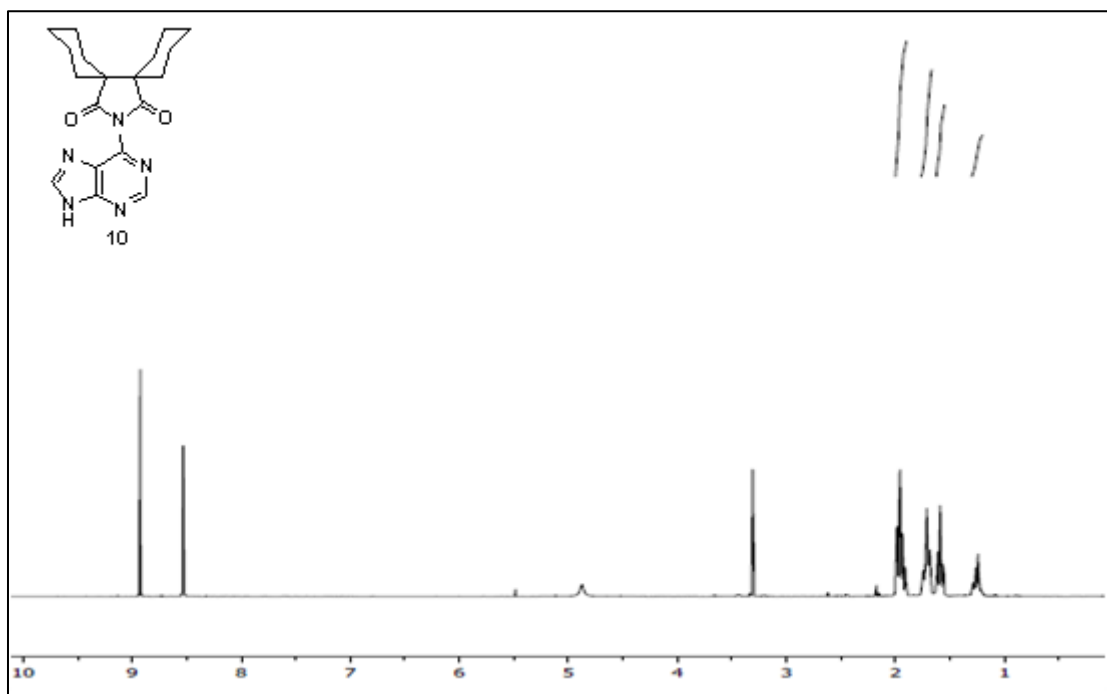
<sup>1</sup>H NMR of Compound 9



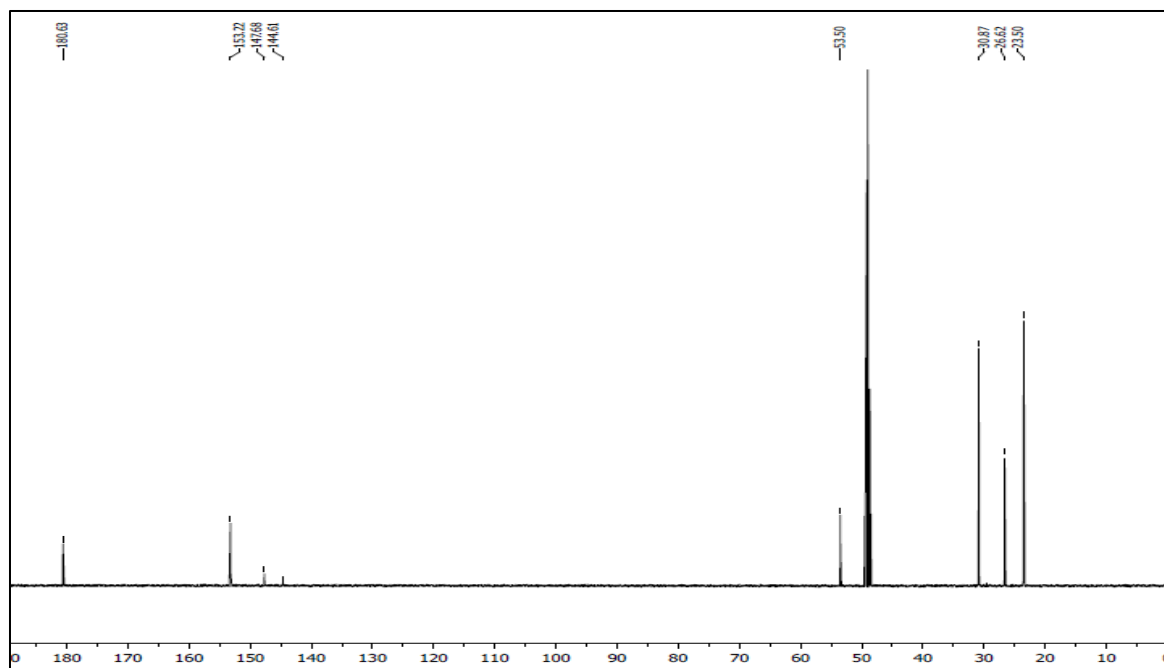
<sup>13</sup>C NMR of Compound 9



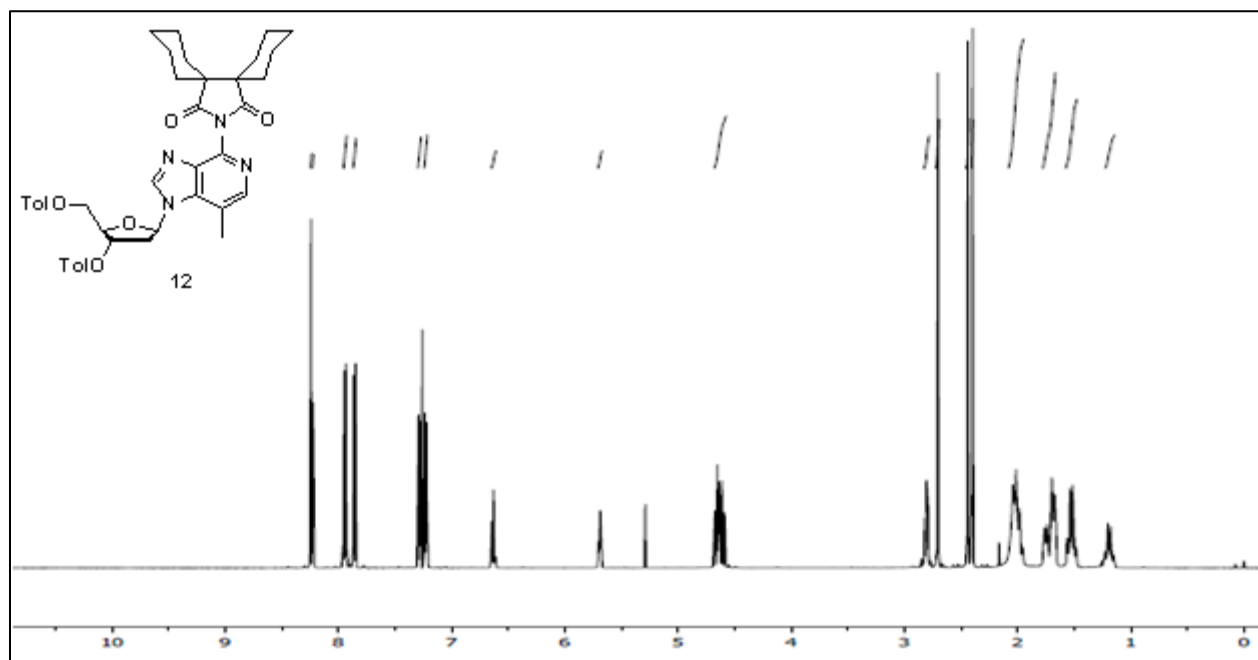
$^1\text{H}$  NMR of Compound **10**



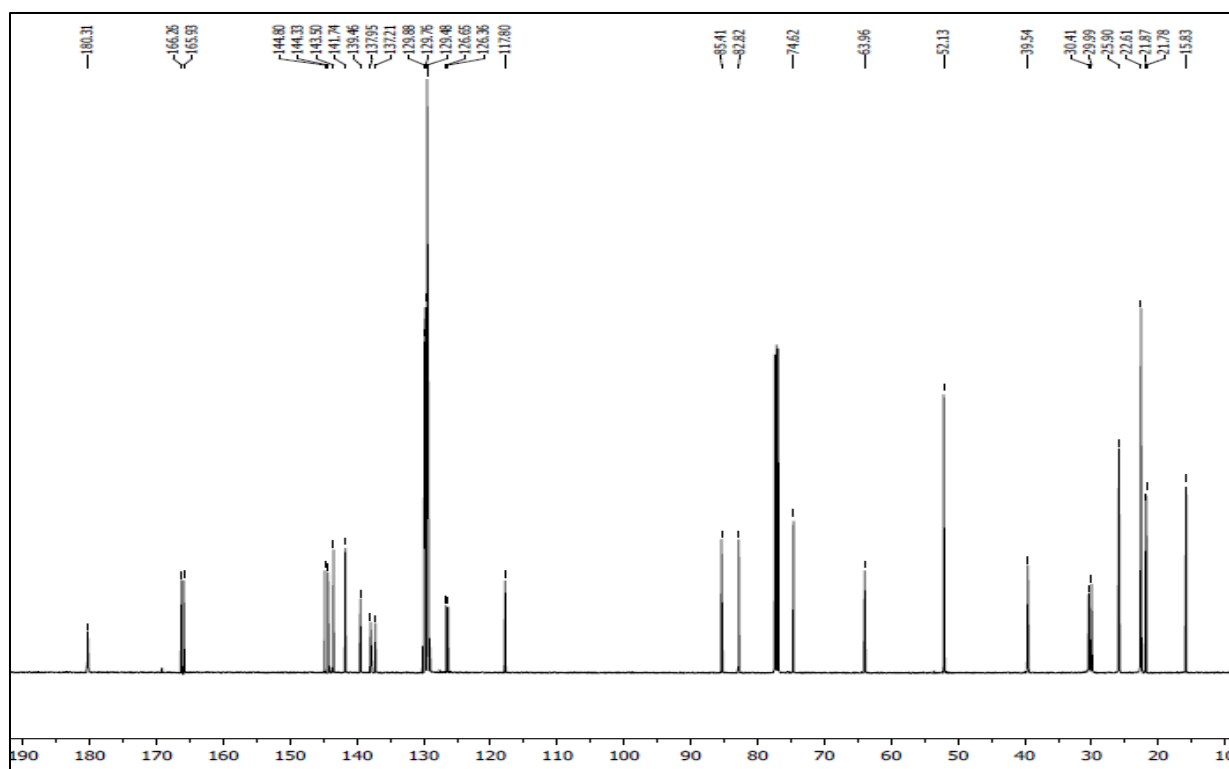
$^{13}\text{C}$  NMR of Compound **10**



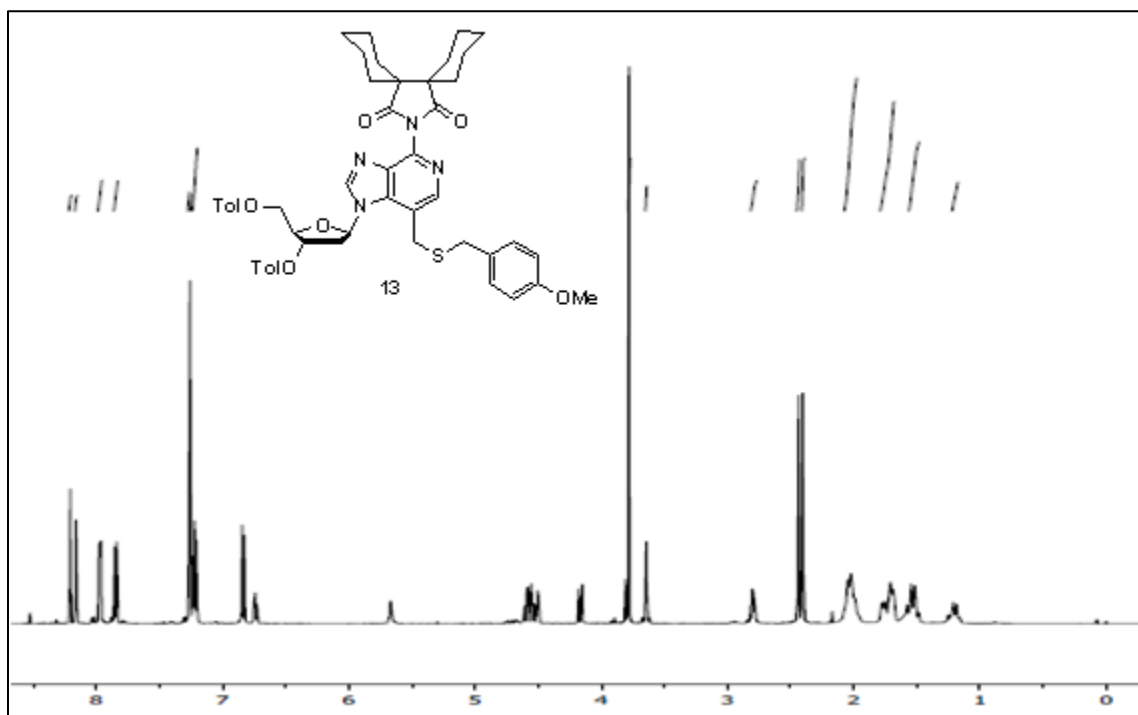
$^1\text{H}$  NMR of Compound **12**



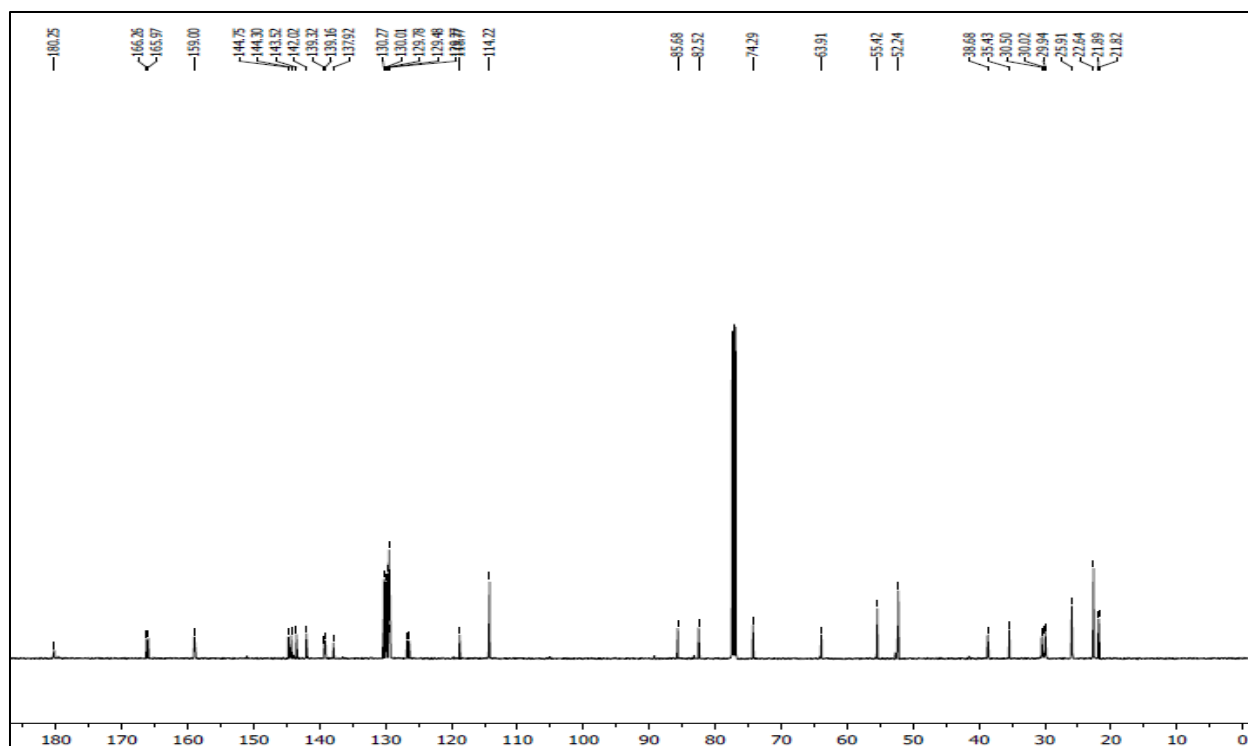
$^{13}\text{C}$  NMR of Compound **12**



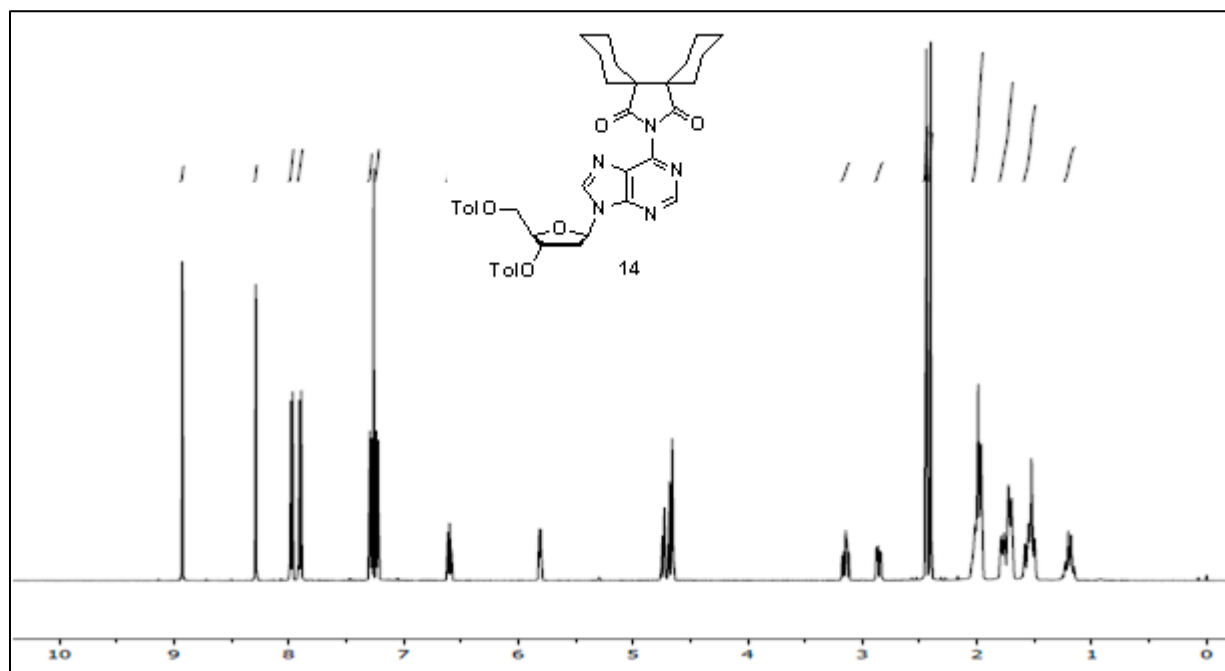
<sup>1</sup>H NMR of Compound **13**



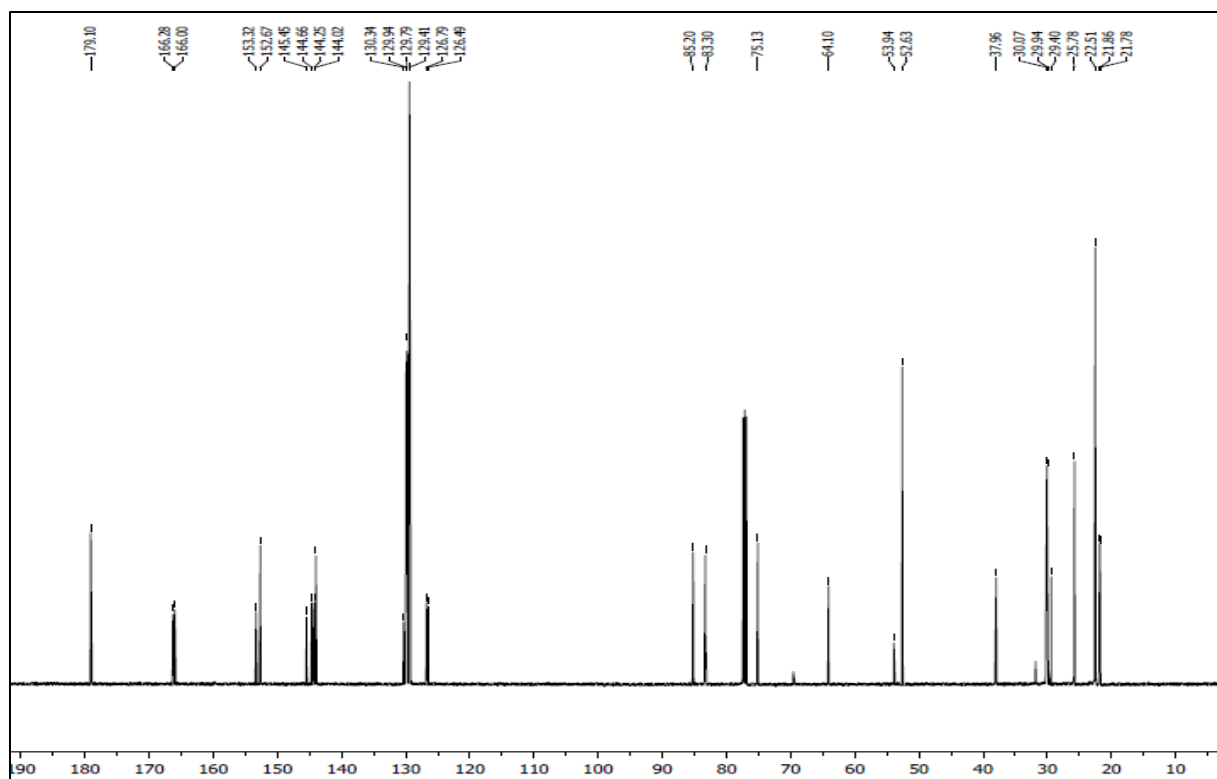
<sup>13</sup>C NMR of Compound **13**



<sup>1</sup>H NMR of Compound **14**

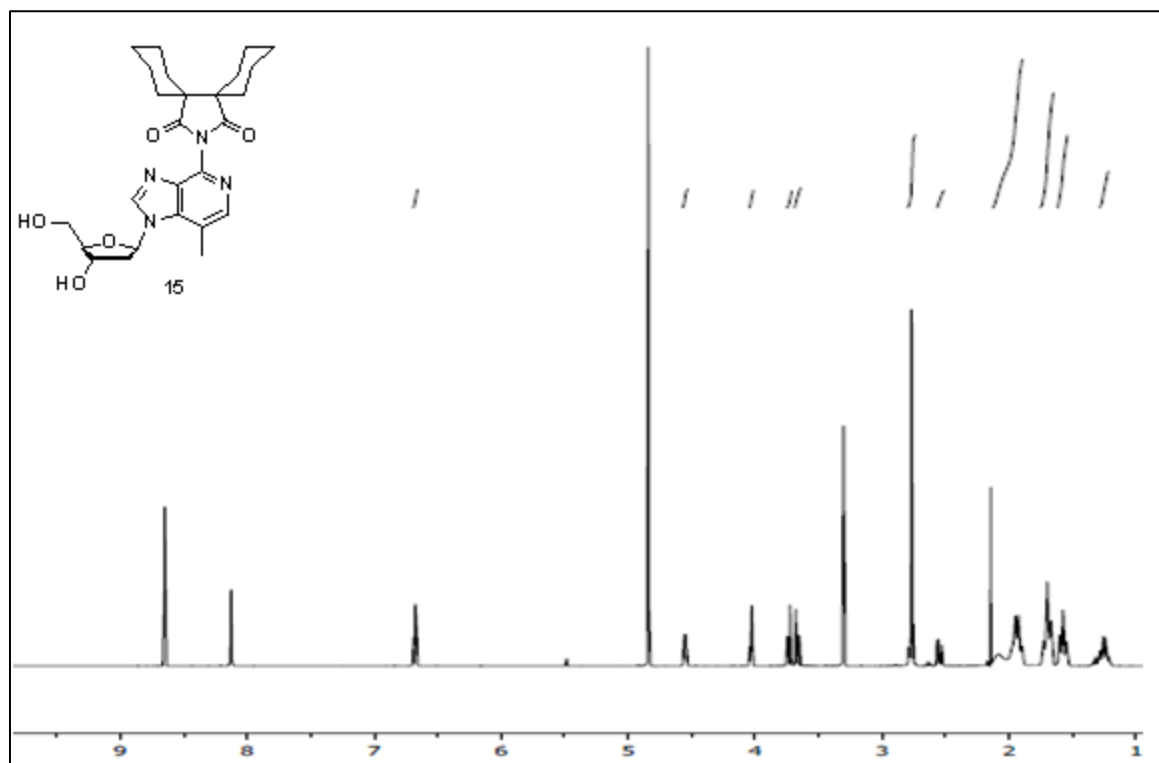


<sup>13</sup>C NMR of Compound **14**

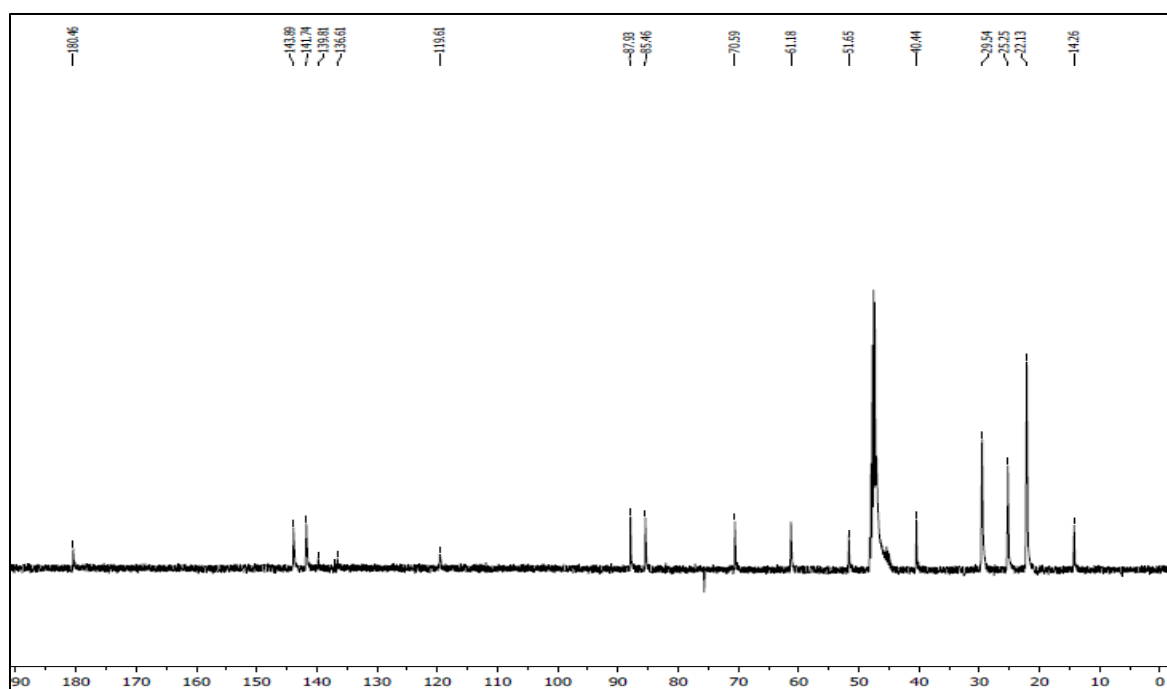




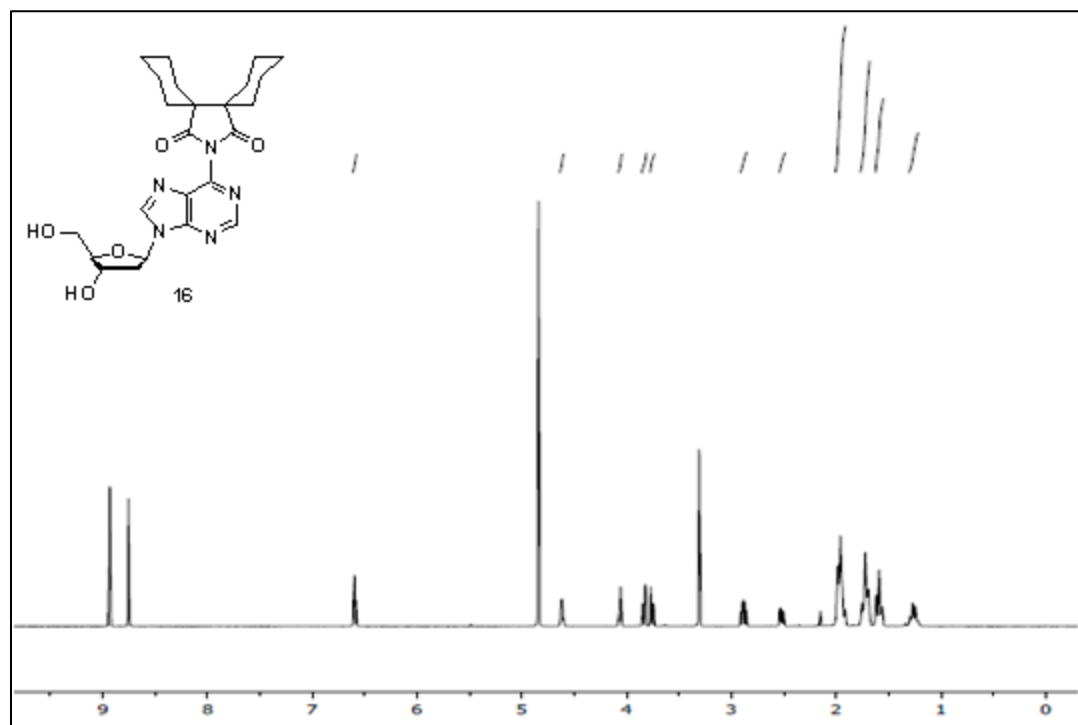
$^1\text{H}$  NMR of Compound **15**



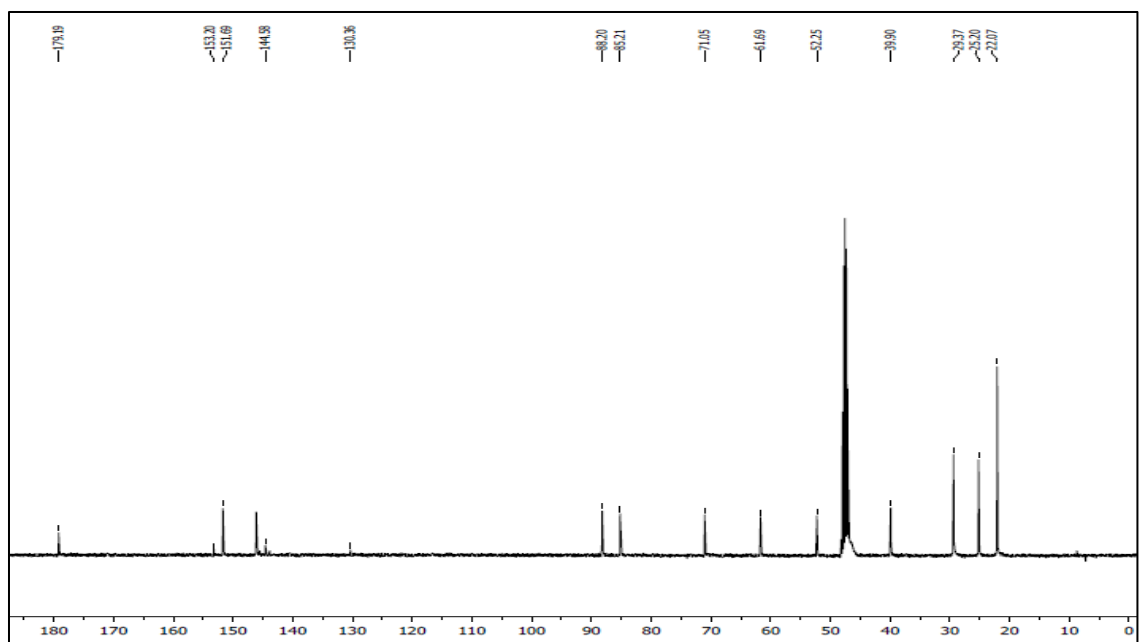
$^{13}\text{C}$  NMR spectrum of Compound **15**



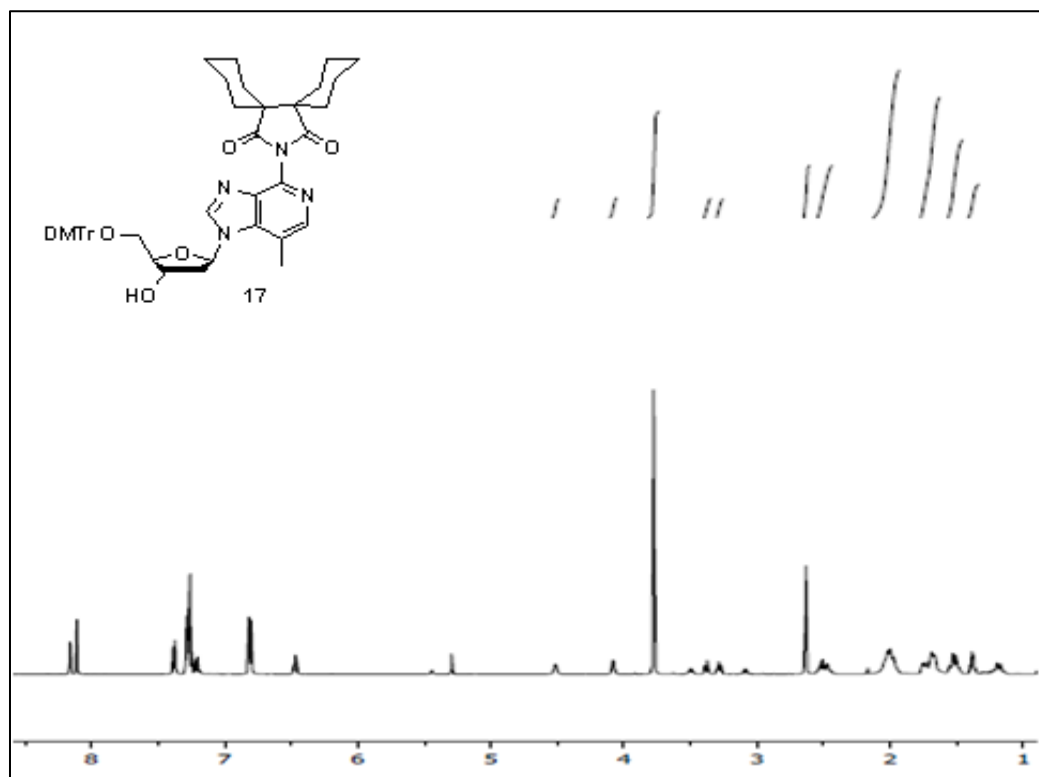
$^1\text{H}$  NMR of Compound **16**



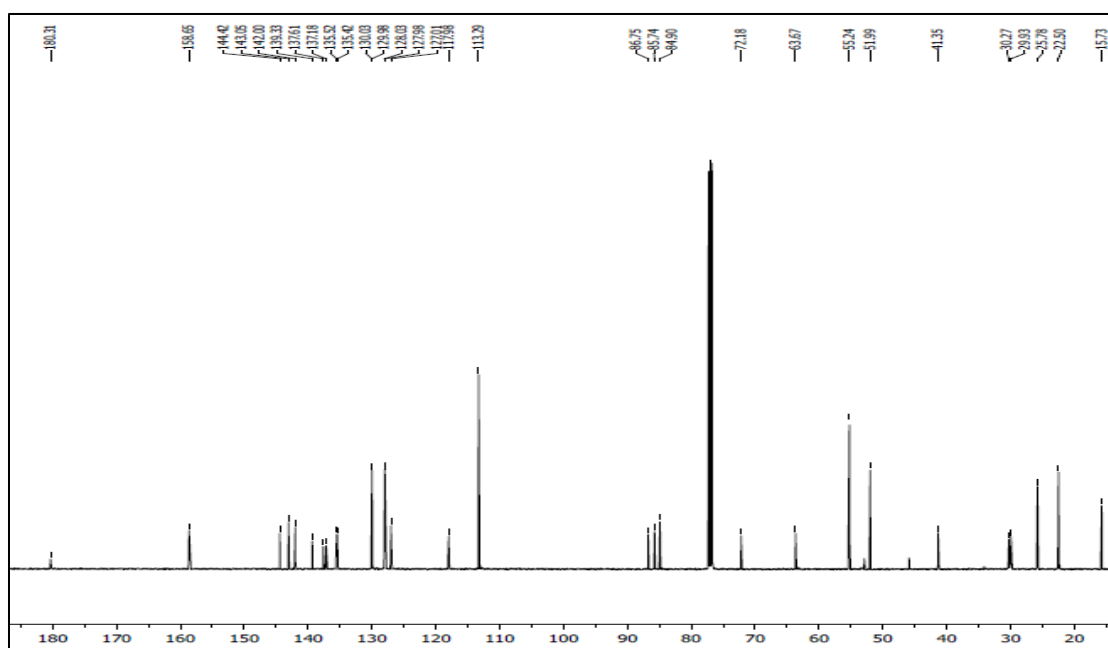
$^{13}\text{C}$  NMR spectrum of Compound **16**



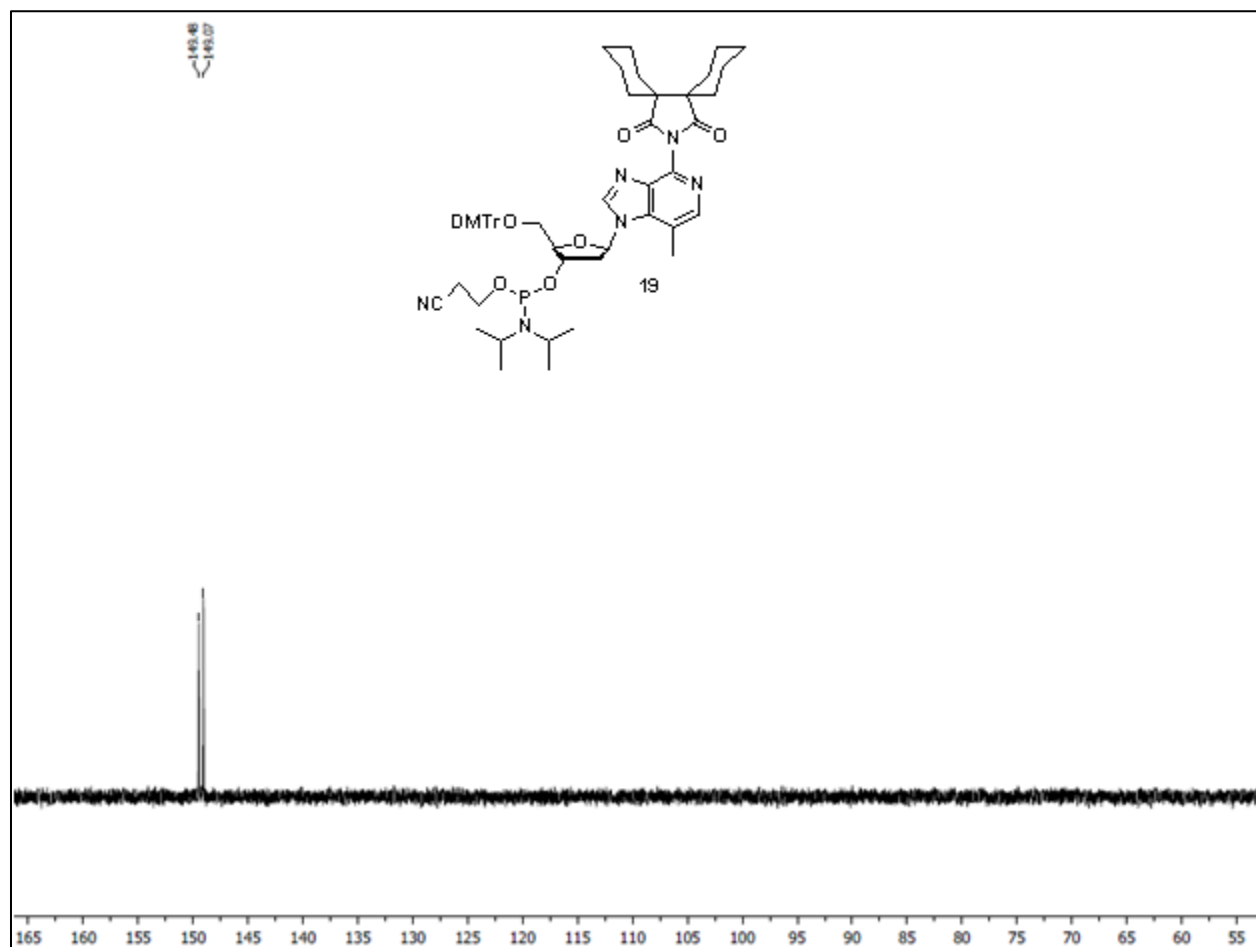
$^1\text{H}$  NMR of Compound **17**



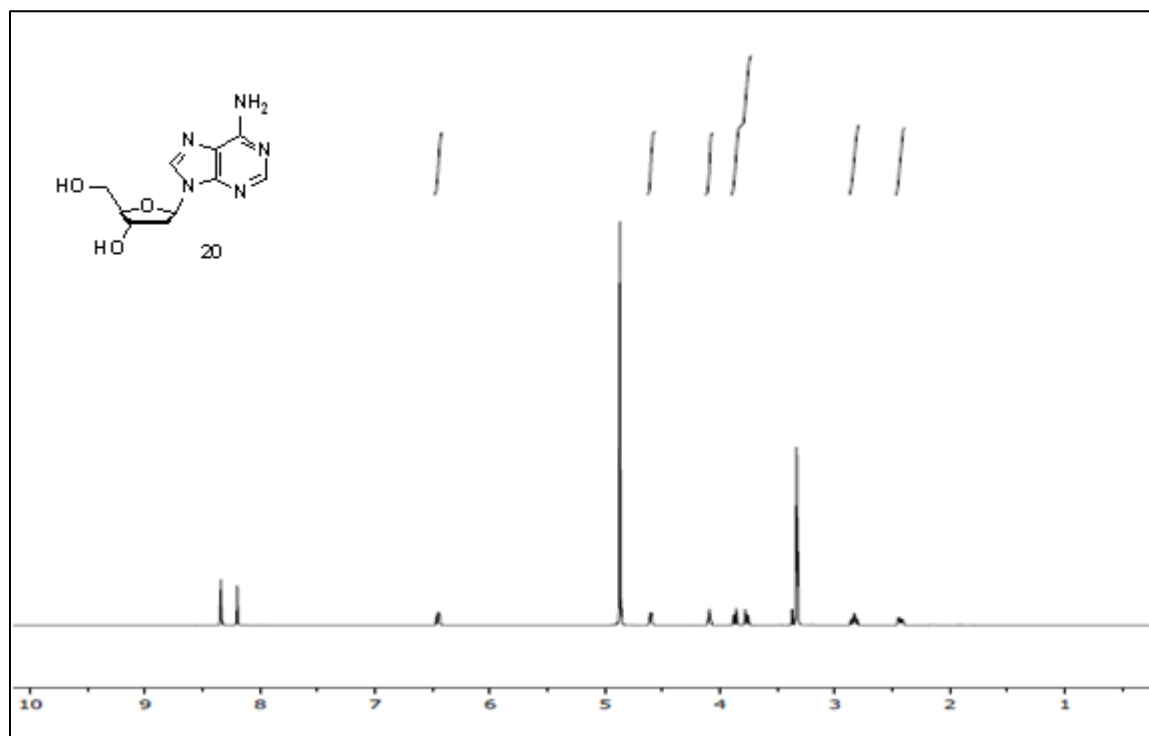
$^{13}\text{C}$  NMR of Compound **17**



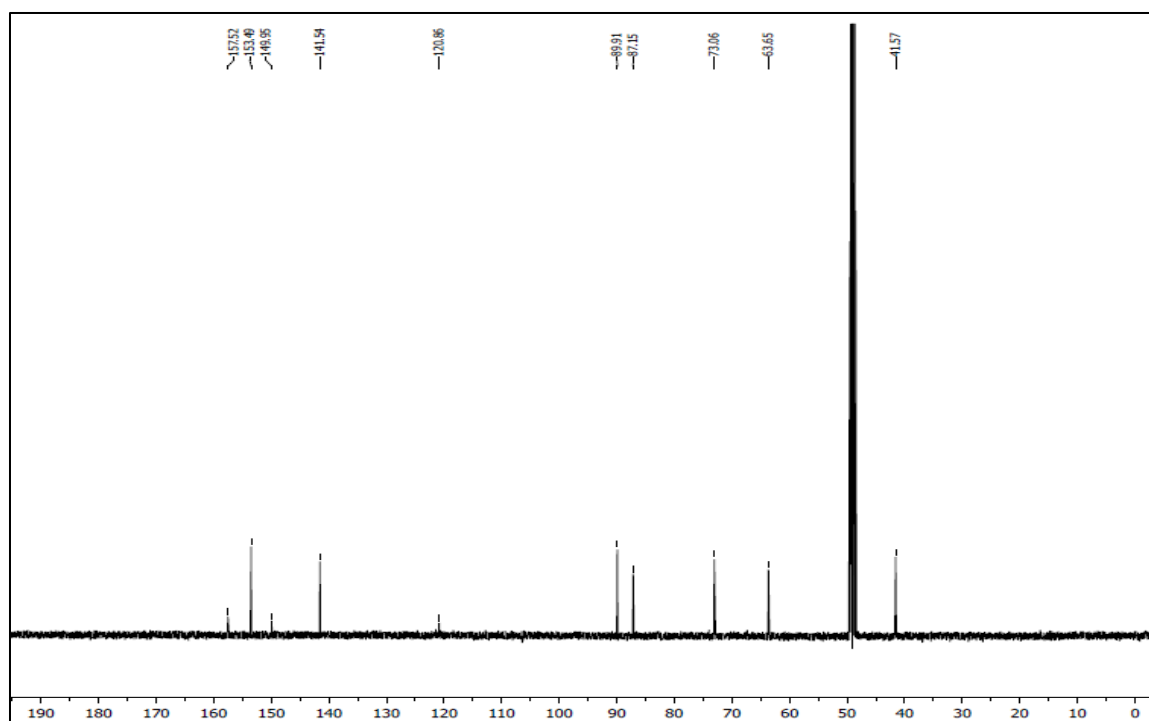
$^{31}\text{P}$  NMR of Compound **19**



$^1\text{H}$  NMR of Compound **20**



$^{13}\text{C}$  NMR of Compound **20**



---

## References

- <sup>1</sup> Beaucage, S. L.; Caruthers, M. H. *Tetrahedron Lett* **1981**, 22, 1859.
- <sup>2</sup> Usman, N.; Ogilvie, K. K.; Jiang, M-Y; Cedergren, R. J. *J. Am. Chem. Soc* **1987**, 109, 7845.
- <sup>3</sup> a) Mansuri, M. M.; Starrett, J. E.; Wos, J. A.; Tortolani, D. R.; Brodfueher, P.R.; Howell, H. G.; Martin, J.C. *J. Org. Chem.* **1989**, 54, 4780. b) Robins, M. J.; Hansske, F.; Low H.N.; Park, J.I. *Tetrahedron Lett.* **1984**, 25, 367.
- <sup>4</sup> Wilson, L. J.; Hager, M. W.; El-Katan, Y. A.; Liotta, D. C. *Synthesis* **1995**, 12, 1465-1479.
- <sup>5</sup> Kazimierczuk, Z.; Cottam, H. B.; Revankar, G. R.; Robins, R. K. *J. Am. Chem. Soc.* **1984**, 106, 6379-632.
- <sup>6</sup> a) Rolland, V.; Kotera, M.; Lhomme, J. *Synthetic Commun.* **1997**, 27, 3505-3511. b) Hoffer, M. *Chem. Ber.* **1960**, 93, 2777-2780.
- <sup>7</sup> Vorbrüggen, H., Ruh-Polenz, C. *Org. React.*, **2000**, 55, 1.
- <sup>8</sup> Kawakami, H.; Matsushita, H.; Naoi, Y.; Itoh, K.; Yoshikoshi, H. *Chem. Lett.* **1989**, 235-239
- <sup>9</sup> Keller, F.; Bunker, J. E.; Brown, L. H. *J. Org. Chem.* **1966**, 31, 3840.
- <sup>10</sup> Yamaguchi, T.; Saneyoshi, M. *Chem. Pharm. Bull.* **1984**, 32, 1441.
- <sup>11</sup> Baud, M. V.; Chavis, C.; Lucas, M.; Imbach, J.-L. *Tet. Lett.* **1990**, 31, 4437.
- <sup>12</sup> Kotick, M. P.; Szantay, C.; Bardos, T. J. *J. Org. Chem.* **1969**, 34, 3806. b) Hubbard, A. J.; Jones, A. S.; Walker, R. T. *Nucleic Acids Res., Symp. Ser. 12* **1984**, 6827.
- <sup>13</sup> Irani, R. J.; SantaLucia, J. Jr. *Nucleosides, Nucleotides & Nucleic Acids* **2002**, 21 (11-12), 737-751.

- 
- <sup>14</sup> Breipohl, G.; Will, D. W.; Peyman, A.; Uhlmann, E. *Tetrahedron* **1997**, *53*, 14671–14686.
- <sup>15</sup> Geen, G. R.; Grinter, T. J.; Kinsey, P. M.; Jarvest, R. L. *Tetrahedron* **1990**, *46*, 6903–6914
- <sup>16</sup> a) Minghong Zhong, Ireneusz Nowak, and Morris J. Robins *J. Org. Chem.* **2006**, *71*, 7773–7779. b) Minghong Zhong, Ireneusz Nowak, John F. Cannon, and Morris J. Robins *J. Org. Chem.* **2006**, *71* (11), pp 4216–4221. c) Ireneusz Nowak and Morris J. Robins *J. Org. Chem.* **2006**, *71* (23), 8876–8883.
- <sup>17</sup> Dzolic, Z.; Kristafor, V.; Cetina, M.; Nagl, A.; Hergold-Pavelic, J.; Balzarini, J.; De Clercq, E.; Mintas, M. *Nucleosides, Nucleotides, and Nucleic Acids* **2003**, *22*, 373–389.
- <sup>18</sup> Alarcon, K.; Martelli, A.; Demeunynck, M.; Lhomme, J. *Tet. Lett.* **2000**, *41*, 7211–7215.
- <sup>19</sup> Zhong, M.; Robins, M. J. *J. Org. Chem.* **2006**, *71*, 8901–8906.
- <sup>20</sup> Zhong, M.; Nowak, I.; Robins, M. J. *J. Org. Chem.* **2006**, *71*, 7773–7779.
- <sup>21</sup> Hall, H. K., Jr. *J. Am. Chem. Soc.* **1958**, *80*, 6412–6420.
- <sup>22</sup> Arico, J. W.; Calhoun, A. K.; Salandria, K. J.; McLaughlin, L. W. *Org. Lett.* **2010**, *12*, 120–122.
- <sup>23</sup> Overberger, C. G.; Bilech, H.; Huang, P. T.; Blatter, H. M. *J. Org. Chem.* **1955**, *20*, 1717–1720.
- <sup>24</sup> a) Ritzmann, G.; Klein, R. S.; Hollenberg, D. H.; Fox, J. J. *Carbohydr. Res.* **1975**, *39*, 227. b) Howell, H. G.; Brodfuehrer, P. R.; Brundlage, S.P.; Benigni, D. A.; Sapino, C. J. *Org. Chem.* **1988**, *53*, 85.

---

<sup>25</sup> Salandria, K. J.; Arico, J. W.; Calhoun, A. K.; McLaughlin, L. W. *J. Am. Chem. Soc.*, **2011**, *133*, 1766–1768.



# **Chapter 3**

## **Cyclic Oligonucleotides**

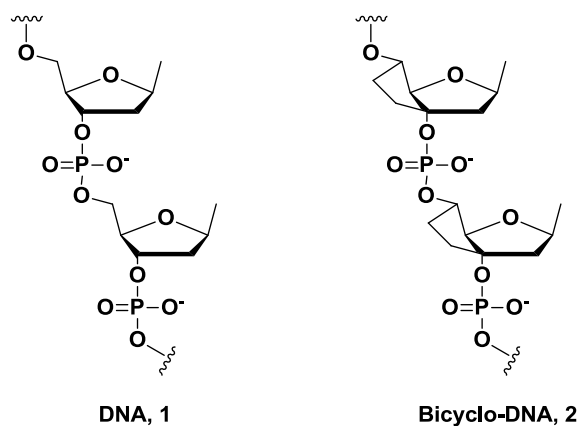
## Background

As we have discussed in section 1.6, the preorganization of modified oligonucleotides is an important aspect for sequence specific recognition of a single-stranded nucleic acid. One of the strategies to rigidify the nucleotide units is by locking the sugar moiety in a fixed conformation by introducing new covalent bonds. Some of the successful candidates are discussed below.

### 3.1. Rigidified Nucleic Acids

#### 3.1.1. Bicyclo-DNA

In this approach the preorganization of DNA is achieved by rigidifying the normally flexible furanose ring by chemical modification. This oligonucleotide analogue has been termed “bicyclo-DNA”<sup>1</sup>. The modification is done by the addition of an ethylene bridge from C-3' to C-5', adding a second five-membered ring to the natural structure (**2**) (Figure 3.1).

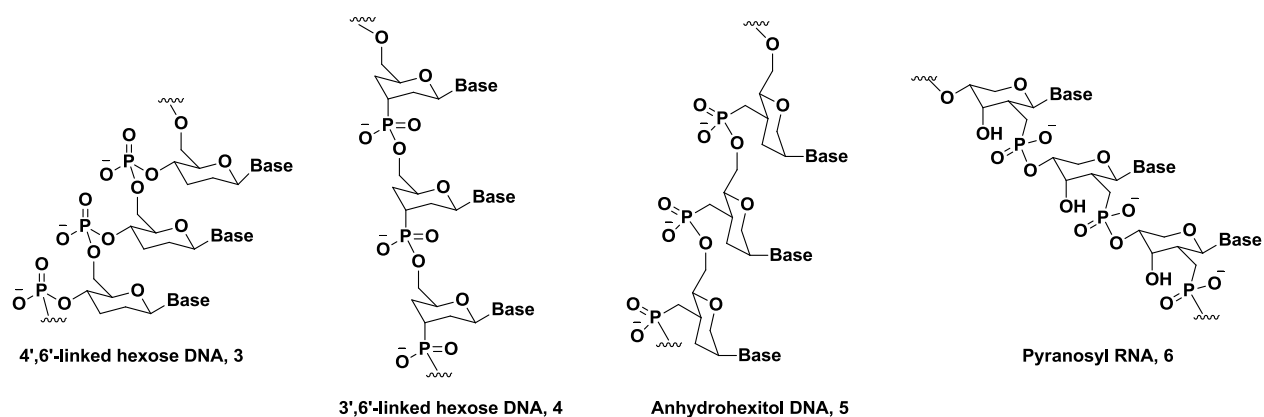


**Figure 3.1:** Illustration of the structure of Bicyclo-DNA compared to DNA.

This analogue has been found in some cases to hybridize strongly with an RNA strand; for example, a strand of bicyclo-DNA with adenine bases binds poly(U) more tightly than does a similar strand of natural adenine-containing DNA, and similar effects have been observed in triplex formation. Some sequences of bicyclo-DNA bind a single-stranded or duplex complement with lower affinity than natural DNA; in part this may be due to the adoption of an unfavorable ring conformation<sup>1</sup>.

### 3.1.2. Hexose-DNAs and RNAs

Six-membered rings are known to be more rigidified than five-membered rings<sup>2</sup>. Furanose sugars have a low barrier to inter-conversion of ring conformers, while pyranoses are generally much more stabilized in a chair conformation. Several analogues of DNA and RNA of this class have been designed and synthesized in which the furanose ring is expanded to a six-membered ring (Figure 3.2)<sup>2</sup>.

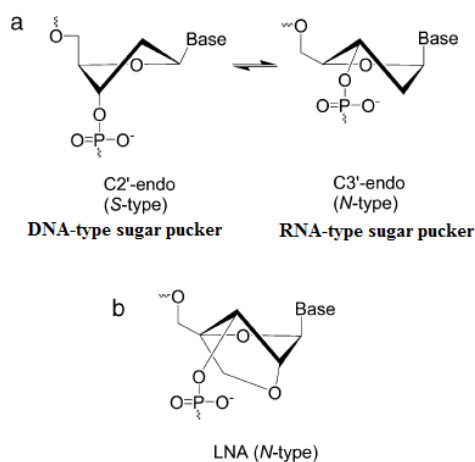


**Figure 3.2:** Examples of Hexose-DNA and RNA.

One fascinating example of this concept is the pyranosyl-RNA analogue (pRNA) (**6**) of Eschenmoser. This structure contains the same number of carbons as RNA but with six-membered rings rather than five-membered ones, and with one fewer freely rotating internucleotide bond (Figure 3.2). Hybridization studies have shown that oligonucleotides constructed with this backbone bind to RNA complementary strands very strongly; this result is attributed in large part to the conformational rigidity of the modified sugar-phosphate backbone<sup>2</sup>.

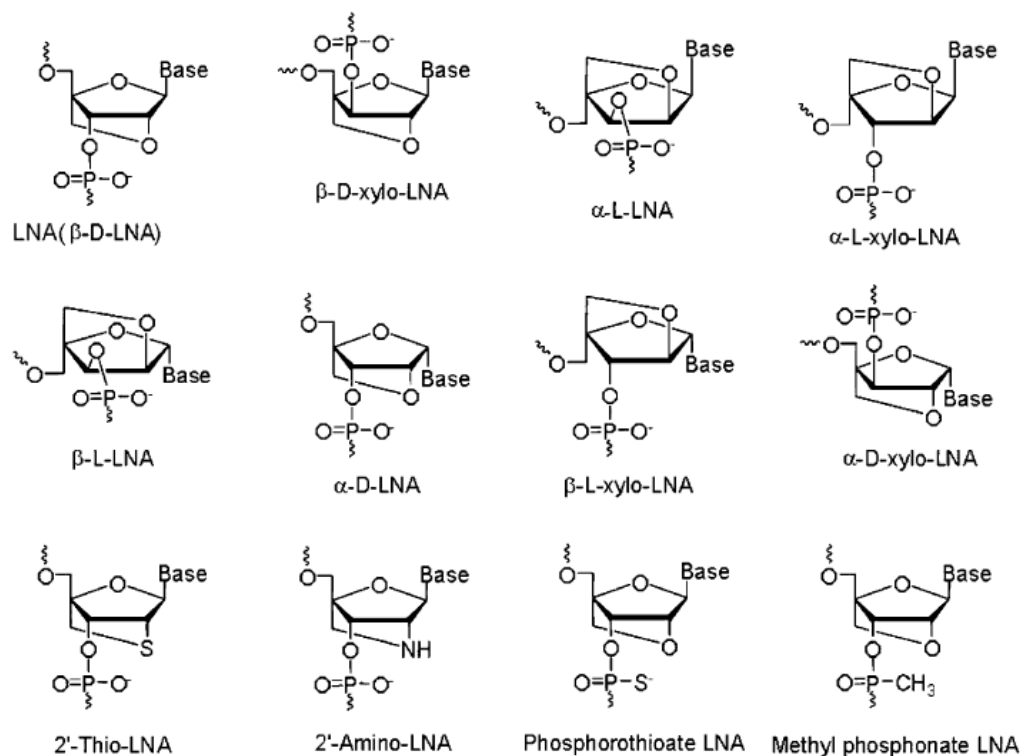
### 3.1.3. Locked Nucleic Acids (LNA)

Locked nucleic acids (LNA) were introduced in 1998 by Wengel and co-workers in order to generate monomers which will favorably lock the sugar in the C3'-*endo* conformation<sup>3</sup>. In this modification, the C2'-oxygen of ribonucleosides is tethered to the C4' via a methylene group. The modification rigidifies the sugar pucker permanently to C3'-*endo* conformation as shown in Figure 3.3b. Subsequently various analogues of these original LNAs have been synthesized (Figure 3.4)<sup>4</sup>.



**Figure 3.3.** Conformation of sugar pucker. a) In DNA and RNA; b) In LNA.

LNA is a unique class of rigidified oligonucleotide because of their remarkable desirable binding affinity<sup>5</sup>, stability in presence of nucleases<sup>6</sup> and serum<sup>7</sup> and low toxicity<sup>8</sup>.

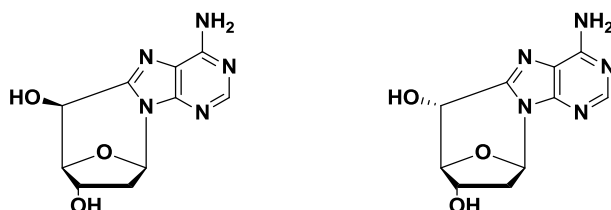


**Figure 3.4.** Selected Examples of LNAs.

### 3.1.4. Cyclic Oligonucleotides

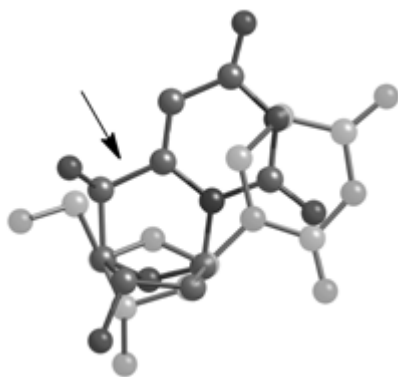
Cyclic oligonucleotides are another class of rigidified oligonucleotides. The main feature of this class is that the nucleobase is locked in the anti-conformation with respect to the sugar by tethering the C5' of the sugar with the C8 of the purine or C6 of the pyrimidine. Broadly, cyclic nucleosides can be categorized in two sub-categories. First, where the sugar and the base are directly connected involving the mentioned atoms<sup>9</sup>. Second, where they are connected via extended chains. We will designate the first and second sub-category as Generation 1 and Generation 2 respectively.

Some of the Generation 1 analogues like 8,5'-cyclopurine represent a unique class of naturally occurring, helix-distorting DNA lesions that result in a covalent linkage between the sugar and base moieties of a single nucleoside, most commonly at 2'-deoxypurines<sup>10</sup>.



**Figure 3.5.** Example of Generation 1: Two epimers of 8,5-cyclo-2'-deoxyadenosine.

Our group has studied extensively the binding, selectivity and affinity of the Generation 1 containing oligonucleotides. Analogues of diastereomerically pure dA (Figure 3.5), dU, U and dG have been synthesized and incorporated in the DNA. The modified strands were used to bind native DNA, RNA and modified DNA strands. We will discuss the dG analogue in section 3.2. In all the cases, the  $T_m$  and the change in hyperchromicity were lower than the native DNA duplex control. Our investigation has proven that the linkage between C5' and C8 “pulls” the base back from its native Watson and Crick bonding position and create an unfavorable  $\chi$  angle, leading to non-effective hydrogen bonding with its corresponding base partner on the complimentary strand (Figure 3.6).



**Figure 3.6:** Overlay of dU and FdU

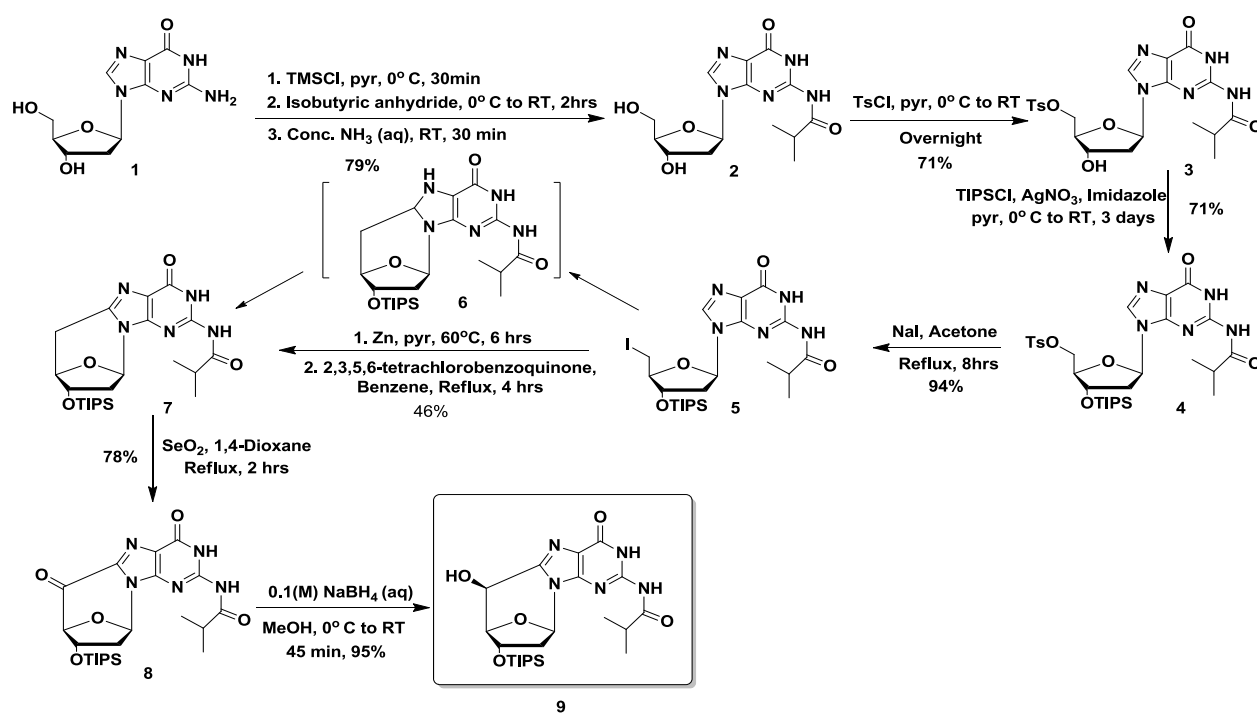
With these results, we decided to develop the Generation 2 cyclic-nucleotides. My colleague, Yueh Han, and I collaborated to synthesize 6,8'-cyclo-2'-deoxyadenosine. Our goal with 6,8'-cyclo-2'-deoxyadenosine is to connect C5' and C8 via a methylene group. It should be able to address two issues- i) we could effectively “push” back the base closer to the original position; and ii) impart some degree of flexibility in the linkage so that the base can adopt a favorable  $\chi$  angle. Both of these should lead to better preorganization of these nucleosides. Details of the synthesis will be discussed in section 3.3.

### **3.2. Synthesis of 8,5-(*S*)-cyclo-2'-deoxyguanosine\***

The synthesis starts with readily available 2'-deoxyguanosine (Scheme 1). The 2-amino group is first protected with standard isobutyric group (**2**). This is followed by the selective installation of tosyl-group at 5'-hydroxyl group (**3**). This step is fairly selective for 5'-hydroxyl over 3'-hydroxyl because the former is primary alcohol and preferentially react first. A trace amount of di-protected product was observed. In the next step the free 3'-hydroxyl was protected with TIPS-group (**4**). This reaction was found to be extremely slow. Attempts were made to heat the reaction mixture. However it was found that heating makes the reaction messy. It was also observed that without AgNO<sub>3</sub> the reaction was even slower.

---

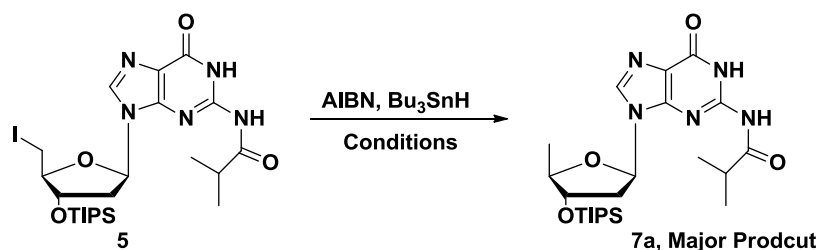
\*The compound numbers are reset here for the rest of Chapter 3. All the figures, schemes and references are continued.



**Scheme 3.1:** Route to synthesize 8,5-(*S*)-cyclo-2'-deoxyguanosine

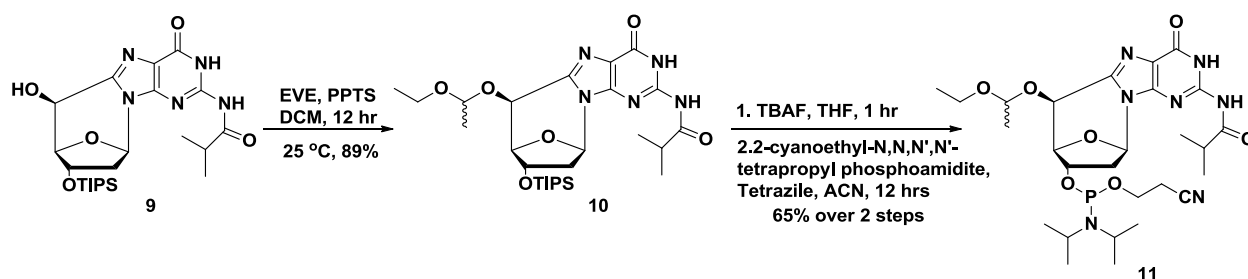
In the next step, an iodo-group (**5**) was installed at C5' by an S<sub>N</sub>2 reaction. This is a very high yielding step with no other side products. The next step is the cyclization reaction by heating the iodo-compound (**5**) with zinc in pyridine to give compound **6**. To reinstall the aromaticity, compound **6** was oxidized by a reducing agent, tetrachloroquinone, to give compound **7**. The overall yield of these two steps was 46%. It was not great; but it was reproducible. In another attempt, a combination of AIBN and Bu<sub>3</sub>SnH was used. This radical based methodology was found to be inefficient. A variety of conditions were tried. The maximum yield of desired product that was observed was only 19%. This major product of this reaction was compound **7a**.





**Scheme 3.2:** Failed attempt to cyclize 5 with AIBN and  $\text{Bu}_3\text{SnH}$ .

We needed the hydroxyl group back at the C5' position. To do that, first a keto-group was introduced (**8**) at that position by selenium dioxide in reasonable yields. Reduction of keto-compound with  $\text{NaBH}_4$  produced (*S*)-epimer exclusively. Nucleophilic attack by a charged nucleophile, in this case a hydride anion, on the unsaturated trigonal center of the carbonyl of a ketone occurs following the Bürgi–Dunitz angle of around  $107^\circ$ . The reason of the high selectivity is the steric hindrance of the sugar moiety that discourages the attack of the hydride from the top face of the six-membered ring according to the Bürgi–Dunitz angle. However, the alternative attack of the hydride from the bottom face is unhindered. This leads to the (*S*)-epimer (**9**) exclusively.



**Scheme 3.3:** Synthesis of the phosphoramidite 11.

To incorporate compound **9** in the modified DNA sequence, we need to prepare its corresponding phosphoramidite (**11**). In general the requirement is to install a DMTr-group at C5'-hydroxyl and phosphoramidite at the C3'-hydroxyl. Numerous attempts were made to install the DMTr-group at C5'-hydroxyl. It resulted in an insignificant

amount of the desired product. Either the starting material was recovered or, in some instances heating the reaction mixture led to the decomposition of the starting material. It was realized that after the cyclization, the rigidity of the molecule increases the bulk in the vicinity of the reaction center, C5'-position. DMTr-group itself is a bulky group. So the spatial steric hindrance might be the reason of the failure of the step.

Alternatively, we needed a group that does not have a huge steric requirement but should be cleaved during the DMTr-deprotection step of the automated DNA synthesis. It was found that ethyl vinyl ether (EVE) is a suitable candidate. When EVE was allowed to react with compound **9**, it gave two diastereomers (**10**) in an excellent overall yield. The two diastereomers co-migrated on TLC and silica gel column chromatography. It was virtually impossible to separate them. The single molecular peak in high resolution mass spectrometry confirmed that the product was clean. The proton NMR confirmed the presence of the two diastereomers. Since these diastereomers would not affect the DNA synthesis, the mixture of products was used for the next step. In the next step, the TIPS group was removed using TBAF, followed by installation of phosphoramidite group at the C3'-hydroxyl (**11**). The presence of the two phosphorus diastereomers and the two EVE diastereomers could be confirmed by the presence of four phosphorus resonances in the final product. It is now ready to be incorporated in DNA sequence.

With this phosphoamidite in hand we prepared three 12 nucleotides long self-complementary sequences (Entry 2-4, Table 3.1). The corresponding native DNA strand (Entry 1) was also prepared. Before we study the binding properties of these modified

strands, we need to make sure that they get incorporated in the DNA strands. This was challenging because we have replaced the widely used DMTr-group by EVE-group at the C5' position for the cyclo-dG derivative. The DNAs were characterized by MALDI-mass spectrometry. The results are tabulated in Table 3.1.

**Table 3.1.** Sequences and Molecular Masses (G\* = Modified dG)

Entry	Sequence (5'→3')	Calculated Mass	Found Mass	Comments
1	d(GCTCACGTGAGC)	3645.7	3646.1	Confirmed.
2	d(G*CTCACGTGAGC)	3643.6	3643.9	Confirmed, although some without cyclo-dG is seen (3316.5).
3	d(GCTCACGTG*AGC)	3643.6	3644.42	Confirmed, but with a large amount of minus cyclo-dG (3316.7).
4	d(GCTCACG*TGAGC)	3643.6	3644.5	Confirmed, but with a large amount of minus cyclo dG (3316.7).

The native sequence was pure (Entry 1). But the results of the DNAs containing the modification were a little puzzling. All of them showed we had the required strand. However, the samples also contained a strand containing 11 nucleotides. The modified nucleotide was missing. Entry 2 can be rationalized from the fact that the HPLC trace of product had a mild shoulder that could not be completely resolved. That explains why we had some strands without the modification. For entry 3 and 4, the modified nucleotide is within the sequence and the HPLC traces looked clean. But still DNA strands without the modification were observed. The only reason it might happen is if the coupling efficiency of the modification was reasonably poor. As a result, there were growing DNA strands without the modification that were not completely capped during the capping step of the

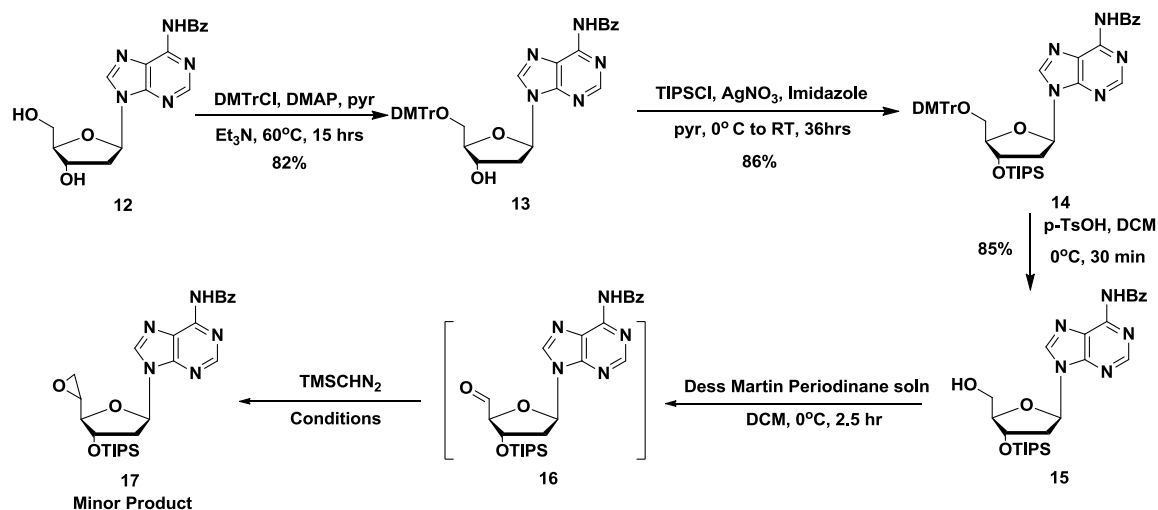
automated DNA synthesis. Hence they kept on growing, resulting in DNA strands with 11 residues. The HPLC purification might not be efficient enough to resolve the difference of one nucleotide. The only way to confirm this would be to perform gel purification. This is currently being pursued. With the pure strands it would be possible to perform further studies.

### 3.3. Synthesis of 6,8'-cyclo-2'-deoxyadenosine

The previous report closest to our molecules was made by Ueda *et al*<sup>11</sup>. They synthesized the ribose analogue, 6,8'-cycloadenosine. Not too many biophysical or biochemical studies were performed with the oligonucleotides containing 6,8'-cycloadenosine. Neither was the issue of stereochemistry at the newly generated C5' chiral explored. One of the reasons might have been the difficulty to synthesize this molecule in reasonable quantity to investigate their properties. Inspired from this and the limitations from the Generation 1 cyclo-nucleosides, we decided to develop a route to synthesize 6,8'-cyclo-2'-deoxyadenosine to investigate their biophysical properties.

The synthesis starts with 2'-deoxyadenosine (Scheme 4). The multi-step one-pot procedure to protect the C6-amino group (**12**) is followed by protection of 5'-OH and 3'-OH by DMTr and TIPS groups (**13** and **14**) respectively. It was followed by the regeneration of 3'-hydroxyl by the removal of DMTr-group (**15**). The next step was to oxidize the 5'-OH to aldehyde (**16**). A variety of conditions were tried including Swern oxidation<sup>12</sup>, Ley oxidation (using TPAP and NMO)<sup>13</sup>, PCC oxidation<sup>14</sup> and Dess-Martin oxidation<sup>15</sup>. It was observed that the Dess-Martin oxidation gave the best results. It was a

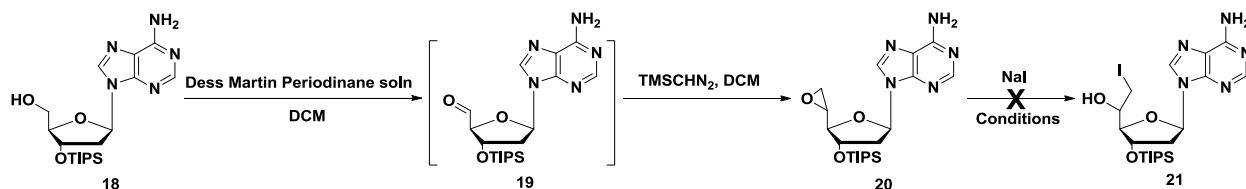
clean, quick and efficient method for controlled oxidation of the primary hydroxyl to aldehyde. To introduce the extra carbon at C5', the aldehyde was allowed to react with trimethylsilyl diazomethane to generate the epoxide (**17**).



**Scheme 3.4.** Attempt to synthesize C5'-C6' epoxide.

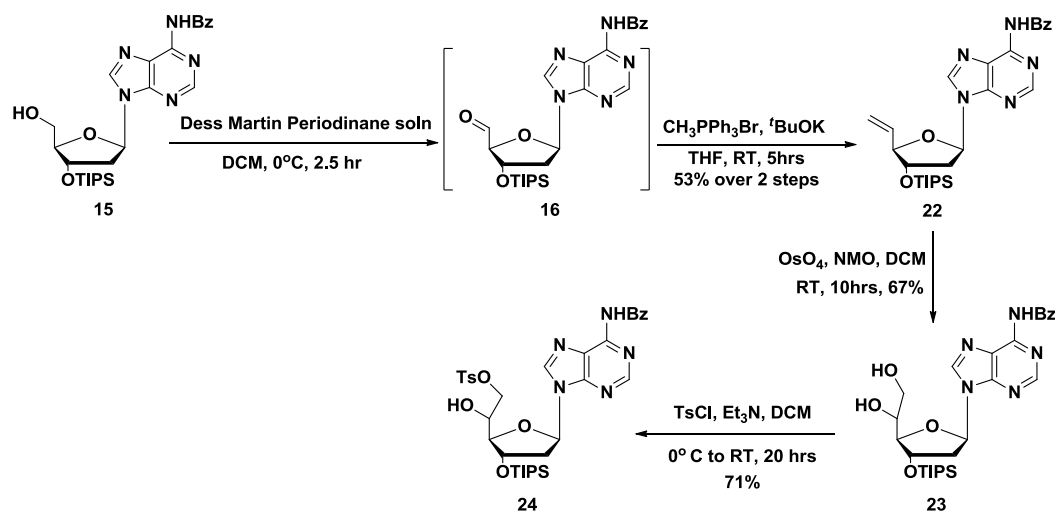
After trying different condition, only trace amounts of the desired product (**17**) were obtained. The by-product could not be characterized as the reaction was messy. The only insight we had was that there was a consistent peak in the mass spectrometry that was approximately 14 units more than the desired product. It led us to think, though unlikely, that the C=O moiety of the  $\text{-NHCOPh}$  at the C6 might be reacting. So the benzoyl group was removed from the C6-amino position and it (**18**, Scheme 5) was then subjected to the trimethylsilyl diazomethane. The reaction was still messy and the product could not be isolated. However, the mass spectrometry confirmed the starting material was consumed and the desired product (**20**) was present. The crude mixture was allowed to react with NaI to install the iodo-group at C6'-position (**21**). We always recovered back the starting material. Mildly acidic conditions, which are supposed to open the epoxide in the undesired way, were also employed but we did not see any conversion. These results

were puzzling. The problem could not be solved, so instead we took an alternative route (Scheme 6).



**Scheme 3.5.** Attempt to open the epoxide.

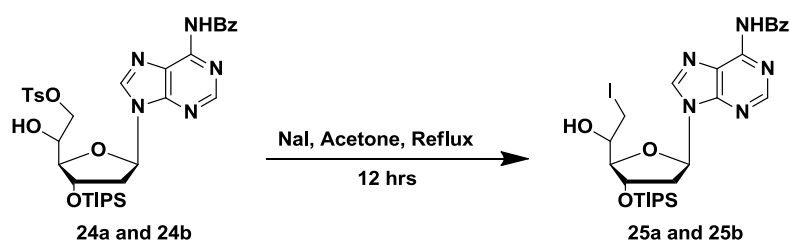
In this route instead of introducing the extra carbon via diazomethane reagent, we decided to use a Wittig reaction. Similar reaction conditions were applied to both **16** and **19**. It was found that the yield from **16** was better. Moreover, we needed the benzoyl group at C6-amino position for DNA synthesis. So we decided to move forward using **16**.



**Scheme 3.6:** Synthesis of precursor to the iodination at C6' position

To regenerate the hydroxyl group at C5'-position, an Upjohn dihydroxylation reaction<sup>16</sup> was performed with **22** to generate **23**. This reaction generated two diastereomers at C5'

position. However, they were inseparable by column chromatography. So the mixture was moved forward according to the scheme. To connect the C6' of the sugar with the C8 of the purine, we needed an iodo-group at C6'-position. For that, the 6'-OH needed to be converted to a better leaving group. It was achieved by reacting **23** with p-TsCl. We obtained the two diastereomers (**24a** and **24b**) and the di-tosylated products in the ratio of 1:1:0.2. Those were separated at this stage. Each of the diastereomers were refluxed in acetone with NaI to derive the desired iodo-compound (**25a** and **25b**).



**Scheme 3.7:** Installing the iodo-group at C6' of both the diastereomers.

The next step is cyclization. Two different methodologies were tried. In the first procedure, the iodo-compound was treated with AIBN and Bu<sub>3</sub>SnH in refluxing condition. This resulted in either no product or very trace amount of product. It was not viable considering it is deep in the synthesis scheme. The second procedure is to reflux the iodo-compound with Zn. Similar conditions resulted in reasonable yields for the 8,5-(*S*)-cyclo-2'-deoxyguanosine discussed in section 3.2. But with **25**, the results are not encouraging yet. I had to stop at this point because I was running out of material. One of my coworkers, Han Yueh, is collaborating with me in this project and trying to optimize the step.

### 3.4. Experimental and Methods

#### *In Reference to section 3.2.*

Compound **4**: The compound **3** (4.9 g, 10 mmol), imidazole (2.03 g, 30 mmol), and silver nitrate (5.50 g, 30 mmol) were dissolved in dry pyridine (100 mL) and cooled to 0 °C. Triisopropylsilyl chloride (6.3 ml, 30 mmol) was added and the reaction was warmed to ambient temperature. After stirring for 3 days the reaction, the starting material was consumed. The reaction was cooled to 0 °C in ice-bath and quenched with water. The solvents were removed under reduced pressure and the residue dissolved in dichloromethane and washed with water three times. The organic layer was dried over sodium sulfate and concentrated under vacuum. The crude product was purified by flash chromatography (1:49 MeOH:DCM) to yield compound **4** (4.56 g, 71%).  $R_f$  = 0.49 (1:49 MeOH:DCM).  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  12.07 (s, 1H), 9.16 (s, 1H), 7.75 (d,  $J$  = 8.3 Hz, 2H), 7.69 (s, 1H), 7.35 (d,  $J$  = 8.3 Hz, 2H), 6.13 (dt,  $J$  = 8.8, 4.4 Hz, 1H), 4.56 – 4.47 (m, 2H), 4.13 (tdd,  $J$  = 9.3, 6.3, 3.4 Hz, 2H), 3.30 – 3.21 (m, 1H), 2.64 (dt,  $J$  = 13.7, 6.9 Hz, 1H), 2.45 (s, 3H), 2.21 (td,  $J$  = 14.1, 6.2 Hz, 1H), 1.24 (dd,  $J$  = 8.3, 4.8 Hz, 3H), 1.20 (d,  $J$  = 6.8 Hz, 3H), 1.10 – 1.04 (m, 3H), 1.04 – 0.94 (m, 18H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  179.16, 155.66, 148.07, 147.41, 145.86, 139.27, 132.08, 130.30, 127.85, 123.17, 86.78, 85.42, 73.41, 69.04, 39.23, 36.42, 21.82, 19.12, 18.96, 18.84, 18.10, 18.01, 17.68, 17.14, 13.72, 12.56, 12.14, 12.03. HRMS (ESI-TOF) calcd for  $[\text{M} + \text{Na}]^+$  670.27011, found: 670.27050.

Compound **5**: Compound **4** (4.25 g, 6.56 mmol) and sodium iodide (3.93 g, 26.24 mmol) were dissolved in dry acetone (65 mL) and refluxed for 8 hours. The solid suspension



was filtered through celite and the filtrate was concentrated under vacuum. The crude was purified by flash chromatography (1.5:98.5 MeOH:DCM) to yield compound **5** as a white foam (3.72 g, 94%).  $R_f = 0.44$  (1:49 MeOH:DCM).  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  12.38 (s, 1H), 10.54 (s, 1H), 7.87 (s, 1H), 6.22 – 6.15 (m, 1H), 4.47 (ddt,  $J = 26.1, 5.6, 2.7$  Hz, 1H), 3.99 – 3.91 (m, 1H), 3.42 – 3.29 (m, 2H), 3.03 – 2.90 (m, 1H), 2.84 (ddt,  $J = 13.7, 10.7, 6.7$  Hz, 1H), 2.36 – 2.27 (m, 1H), 1.20 (d,  $J = 6.8$  Hz, 3H), 1.16 (dt,  $J = 9.4, 4.7$  Hz, 3H), 1.09 – 0.93 (m, 21H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  179.96, 156.06, 148.73, 148.12, 138.19, 121.79, 86.78, 84.95, 75.49, 39.62, 36.06, 19.18, 19.05, 18.82, 18.09, 18.05, 17.68, 17.11, 13.91, 12.58, 12.11, 6.95. HRMS (ESI-TOF) calcd for  $[\text{M} + \text{H}]^+$  604.18160, found: 604.18230.

Compound **7**: Compound **5** (3.5 g, 5.80 mmol) and zinc powder (7.6 g, 116 mmol) were mixed in anhydrous pyridine (125 mL) and heated to 60 °C for 6 hours. The zinc powder was removed by filtration through celite and the filtrate was concentrated under reduced pressure. The residue and tetrachloro-1,4-benzoquinone (1.43 g, 5.80 mmol) were mixed in anhydrous benzene (100 mL) and refluxed for 4 hours. The solvent was then evaporated under vacuum and the crude product was purified by flash chromatography (1:49 MeOH:DCM) to yield the cyclized product as a light yellowish white foam (1.27 g, 46%).  $R_f : 0.29$  (3:97 MeOH:DCM).  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  12.04 (s, 1H), 8.98 (s, 1H), 6.28 (d,  $J = 5.0$  Hz, 1H), 4.71 (t,  $J = 7.5$  Hz, 1H), 4.48 (dd,  $J = 7.0, 3.6$  Hz, 1H), 3.43 (dd,  $J = 17.6, 6.1$  Hz, 1H), 2.95 (d,  $J = 17.5$  Hz, 1H), 2.69 (dq,  $J = 13.4, 6.7$  Hz, 1H), 2.56 (dd,  $J = 13.4, 7.1$  Hz, 1H), 2.25 (ddd,  $J = 16.5, 10.5, 6.2$  Hz, 1H), 1.27 (d,  $J = 3.9$  Hz, 3H), 1.25 (d,  $J = 3.9$  Hz, 3H), 1.10 – 1.05 (m, 3H), 1.05 – 0.99 (m, 18H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  178.69, 155.31, 147.42, 145.91, 142.26, 120.37, 84.66, 83.57, 75.47,

48.31, 36.59, 30.76, 29.82, 19.20, 19.02, 17.99, 17.96, 11.96. HRMS (DART-TOF) calcd for  $[M + H]^+$  476.27096, found: 476.26930.

**Compound 8:** Compound **7** (1.0 g, 2.11 mmol) and selenium dioxide (482 mg, 4.35 mmol) were taken in 1,4-dioxane (250 mL) and refluxed for 2 hour. The reaction mixture was filtered through celite and the solvent removed under vacuum. The product was purified by flash chromatography (1:24 MeOH:DCM) to yield **8** as a reddish white foam (769 mg, 78%).  $R_f$ : 0.21 (1:24 MeOH:DCM).  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  11.98 (s, 1H), 8.39 (s, 1H), 6.48 (d,  $J = 5.4$  Hz, 1H), 4.88 (dd,  $J = 7.0, 2.8$  Hz, 1H), 4.83 (s, 1H), 2.83 (dd,  $J = 13.9, 7.0$  Hz, 1H), 2.68 (dt,  $J = 13.8, 6.9$  Hz, 1H), 2.46 (ddd,  $J = 13.8, 5.6, 2.5$  Hz, 1H), 1.65 (s, 3H), 1.31 (t,  $J = 6.9$  Hz, 6H), 1.20 – 1.12 (m, 3H), 1.11 – 1.05 (m, 18H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  183.08, 178.64, 155.37, 149.63, 145.65, 138.56, 123.09, 92.38, 85.63, 72.86, 44.98, 36.87, 19.06, 18.98, 17.99, 11.97. HRMS (DART-TOF) calcd for  $[M + H]^+$  490.24857, found: 490.24853.

**Compound 9:** Compound **8** (0.50 g, 1.02 mmol) was dissolved in methanol (30 mL) and cooled to 0 °C, 0.1 M sodium borohydride aqueous solution (5.75 mL) was added and the stirring reaction was allowed to warm to ambient temperature. After 45 min the reaction was neutralized with 1(N) hydrochloric acid and the solvents were removed under vacuum. The crude product was purified by flash chromatography (3.5:96.5 MeOH:DCM) to yield **9** as a white powder (0.453 g, 95%).  $R_f$ : 0.27 (3:97 MeOH:DCM).  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  12.24 (s, 1H), 9.87 (s, 1H), 6.34 (d,  $J = 4.9$  Hz, 1H), 5.30 (d,  $J = 6.3$  Hz, 1H), 4.99 (dd,  $J = 7.0, 3.8$  Hz, 1H), 4.77 (d,  $J = 6.4$  Hz, 1H), 2.79 (dt,  $J = 13.8, 6.9$  Hz, 1H), 2.58 (dd,  $J = 13.4, 7.2$  Hz, 1H), 2.30 – 2.22 (m, 1H), 1.25 (t,  $J = 6.1$  Hz, 3H), 1.20 (d,  $J = 6.9$  Hz, 3H), 1.08 (td,  $J = 7.5, 3.5$  Hz, 3H), 1.06 – 0.98 (m, 18H).

$^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  179.26, 155.13, 148.18, 147.25, 145.42, 120.10, 86.48, 85.22, 69.29, 64.07, 46.95, 36.15, 19.09, 18.84, 17.84, 17.83, 11.85. HRMS (DART-TOF) calcd for  $[\text{M} + \text{H}]^+$  492.2642, found: 492.26368.

**Compound 10:** Compound **9** (0.083 g, 0.17 mmol) and pyridinium-p-toluenesulfonate (47 mg, 0.19 mmole) were dissolved in dichloromethane (1.7 mL) followed by the addition of ethyl vinyl ether (0.1 mL, 1.02 mmol) and stirred overnight at ambient temperature. The solvents were removed under vacuum and the crude product was purified by flash chromatography (3:97 MeOH:DCM) to yield **10** as a light yellowish white foam (0.085 g, 89%).  $R_f$  = 0.32 (silica gel, 3:97 MeOH:DCM).  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  12.14 (s, 2H), 9.37 (s, 2H), 6.26 (dd,  $J$  = 6.3, 5.1 Hz, 2H), 5.56 (q,  $J$  = 5.2 Hz, 1H), 5.26 – 5.19 (m, 2H), 5.12 (d,  $J$  = 6.2 Hz, 1H), 4.93 (tdd,  $J$  = 12.1, 7.6, 4.5 Hz, 2H), 4.66 (d,  $J$  = 6.4 Hz, 1H), 4.60 (d,  $J$  = 6.2 Hz, 1H), 3.78 (dq,  $J$  = 9.4, 7.1 Hz, 1H), 3.72 – 3.65 (m, 2H), 3.62 – 3.55 (m, 1H), 2.81 – 2.72 (m, 2H), 2.49 (ddd,  $J$  = 13.1, 7.2, 3.2 Hz, 2H), 2.25 – 2.17 (m, 2H), 1.47 (d,  $J$  = 5.3 Hz, 3H), 1.33 (t,  $J$  = 4.8 Hz, 3H), 1.28 – 1.22 (m, 12H), 1.21 – 1.13 (m, 6H), 1.12 – 0.96 (m, 42H). HRMS (DART-TOF) calcd for  $[\text{M} + \text{H}]^+$  564.32173, found: 564.32045.

**Compound 11:** Tetra-n-butylammonium fluoride (1.0 M solution in THF, 0.29 mL) was added to a solution of **10** (0.075 g, 0.133 mmol) in THF (8.5 mL) at ambient temperature and stirred for one hour. The solvent was removed under reduced pressure. The crude material was dissolved in acetonitrile (3.6 mL). 1H-tetrazole in acetonitrile (0.45 M solution in MeCN, 0.12 mL) was added followed by 2-cyanoethyl N,N,N',N'-tetraisopropyl phosphorodiamidite (36  $\mu\text{L}$ , 0.109 mmol) and the mixture was stirred overnight at ambient temperature. The reaction was quenched with 5% triethylethylamine in

methanol and the solvents were removed under vacuum. The residue was dissolved in DCM and washed with saturated sodium bicarbonate three times and once with brine. The organic layer was dried over sodium sulfate. After removing the sodium sulfate, hexanes were added to effect precipitation. The precipitate was collected by filtration and washed with hexanes several times. The precipitate was purified by flash chromatography (0.1:0.9:99 Et<sub>3</sub>N:MeOH:DCM) to yield compound **11** as a white powder in quantitative yield (53 mg, 65% over two steps). <sup>31</sup>P NMR (202 MHz, CDCl<sub>3</sub>) δ 148.93, 148.85, 148.53, 148.34. HRMS (ESI-TOF) calcd for [M + H]<sup>+</sup> 608.29561, found: 608.29600.

### ***DNA Synthesis***

The native and modified 12-mers (entry 1-4, Table 1) were prepared by solid-phase DNA synthesis and deprotected using standard protocols. The analogue nucleotides could be incorporated into DNA strands with essentially the same coupling efficiency as the common nucleotides using the wait time of 20 min during the coupling cycle. All newly-synthesized oligonucleotides were deprotected and cleaved from the CPG beads by 12–14 hr treatment with concentrated ammonia at 55 °C. The supernatant solutions were then decanted and evaporated under high vacuum, then redissolved in 12-15 mL H<sub>2</sub>O for HPLC purification.

Initial purification of native and modified oligonucleotides was accomplished by HPLC (Oligo R3 reverse-phase C18 column, trityl on), starting with 100% solution A using a linear gradient from 0 to 65% solution B over 20 min [solution A: 1 M TEAA (pH 7) in 5% acetonitrile; solution B: 1 M TEAA (pH 7) in 70% acetonitrile]. The DMT-protected

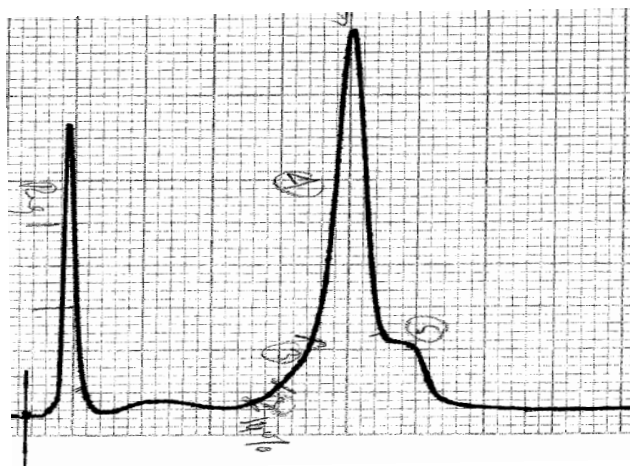
12-mers had retention times of about 17.5 minutes, except for Sequence 1, which had a retention time of 8.5 min due to the presence of a EVE group at the terminal nucleoside (instead of DMTr-group). The collected oligonucleotides were then reduced in volume and detritylated using 80% acetic acid for 30 min at 0 °C. The resulting oligonucleotides were then desalted and lyophilized.

*HPLC traces of DNA (Entry 1, 2, 3 and 4 from Table 3.1) are below.*

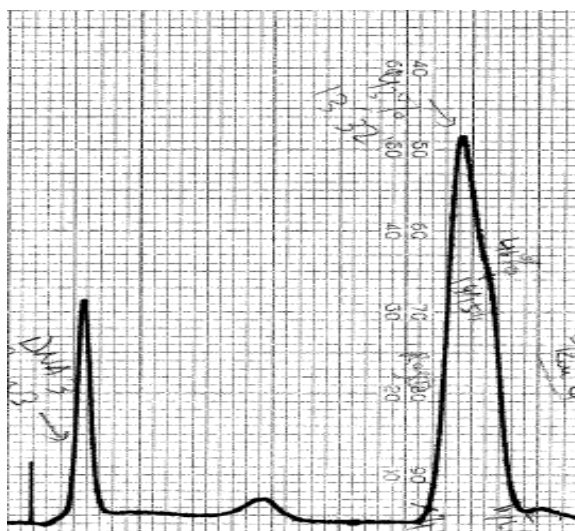
### DNA 1



### DNA 2



### DNA 3



### DNA 4



*In Reference to section 3.3.*

Compound **14**: Compound **13** (9.5 g, 14.44 mmol), imidazole (2.95g, 43.32 mmol) and AgNO<sub>3</sub> (7.36 g, 43.32 mmol) were taken in pyridine (75 mL). The mixture was cooled to 0 °C. To this, TIPSCl (9.2 mL, 43.32 mmol) was added dropwise and the temperature was allowed to warm up to room temperature. The reaction mixture was stirred for 3 days

at ambient temperature. Water was added to quench the reaction followed by removal of the solvent under high vacuum. The crude was dissolved in chloroform and washed with water. The organic layer was collected, dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>, filtered and the filtrate was concentrated in vacuo. It was purified by flash column chromatography (0.1:1.5:98.5 TEA:MeOH:DCM) to yield compound **14** (10.11 g, 86%). *R*<sub>f</sub> = 0.34 (1:49 MeOH:DCM). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 9.32 (s, 1H), 8.72 (s, 1H), 8.21 (s, 1H), 8.03 - 7.99 (m, 2H), 7.57- 7.52 (m, 1H), 7.48 -7.44 (m, 2H), 7.39 (dt, *J* = 8.5, 1.9 Hz, 2H), 7.31 - 7.27 (m, 4H), 7.24 (ddd, *J* = 8.4, 3.2, 1.7 Hz, 2H), 7.20 - 7.17 (m, 1H), 6.81 - 6.76 (m, 4H), 6.50 (dt, *J* = 12.2, 6.1 Hz, 1H), 4.69 (dt, *J* = 5.4, 2.8 Hz, 1H), 4.22 (td, *J* = 4.4, 2.8 Hz, 1H), 3.75 (s, 6H), 3.44 (dd, *J* = 10.4, 4.5 Hz, 1H), 3.32 (dd, *J* = 10.4, 4.4 Hz, 1H), 2.84 (ddd, *J* = 13.0, 7.4, 5.6 Hz, 1H), 2.52 (ddd, *J* = 13.0, 6.0, 2.9 Hz, 1H), 1.08 - 0.99 (m, 21H); <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>) δ 164.72, 158.54, 152.51, 151.52, 149.53, 144.49, 141.65, 135.67, 135.63, 133.73, 132.64, 130.01, 129.99, 128.76, 128.10, 127.87, 127.82, 126.87, 123.59, 113.12, 87.50, 86.56, 85.01, 73.16, 63.52, 55.19, 41.03, 17.98, 12.05. HRMS (ESI-TOF) calcd for [M + H]<sup>+</sup> 814.3994, found: 814.4000.

**Compound 15:** Compound **14** (10.00 g, 12.28 mmol) was dissolved in DCM (120 mL) and cooled to 0 °C. A solution of p-TsOH (5.19 g, 27.02 mmol) in methanol (30 mL) was added dropwise. It was stirred at 0 °C for 30 min. The TLC indicated the complete consumption of compound **2**. The reaction was quenched with satd. aq. NaHCO<sub>3</sub> solution. DCM was added to the solution and was washed by water. The organic layer was collected, dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>, filtered and the filtrate was concentrated *in vacuo*. It was purified by flash column chromatography (3:97 MeOH:DCM) to yield compound **15** (5.34 g, 85%). *R*<sub>f</sub> = 0.24 (1:49 MeOH:DCM). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>)

$\delta$  9.47 (s, 1H), 8.69 (s, 1H), 8.09 (s, 1H), 8.00 - 7.96 (m, 2H), 7.56 - 7.51 (m, 1H), 7.47 - 7.42 (m, 2H), 6.35 (dd,  $J = 9.4, 5.4$  Hz, 1H), 4.75 (d,  $J = 4.8$  Hz, 1H), 4.17 (d,  $J = 0.7$  Hz, 1H), 3.93 (dd,  $J = 12.8, 1.6$  Hz, 1H), 3.74 (d,  $J = 8.6$  Hz, 1H), 3.00 (ddd,  $J = 13.0, 9.4, 5.0$  Hz, 1H), 2.29 (dd,  $J = 12.4, 5.6$  Hz, 1H), 1.11 - 1.01 (m, 21H);  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  164.82, 152.02, 150.78, 150.29, 142.64, 133.53, 132.79, 128.76, 128.00, 124.50, 90.53, 87.67, 74.10, 63.25, 41.76, 17.96, 11.97. HRMS (ESI-TOF) calcd for  $[\text{M} + \text{H}]^+$  512.2688, found: 512.2693.

Compound **16**: Compound **15** (5.25 g, 10.26 mmol) and solid DMP (6.53 g, 15.39 mmol) were taken in DCM (150 mL) and stirred for 2.5 hrs at ambient temperature. The reaction was quenched by a 1:1 mixture of satd. aq. solution of  $\text{Na}_2\text{S}_2\text{O}_3$  and  $\text{NaHCO}_3$ . DCM was added to the solution and was washed by water. The organic layer was collected, dried over anhydrous  $\text{Na}_2\text{SO}_4$ , filtered and the filtrate was concentrated *in vacuo* to yield the crude compound **16**. The crude was used in the next step.

Compound **22**:  $\text{CH}_3\text{PPh}_3\text{Br}$  (11 g, 30.78 mmol) and  $t\text{BuOK}$  (3.45 g, 30.78 mmol) were suspended in THF (150 mL). It was stirred at ambient temperature for an hour. A solution of the crude compound **16**, from previous step, in THF (37 mL) was added slowly to the reaction mixture. It was allowed to stir for another 5 hrs. The reaction was quenched by satd. aq. solution of  $\text{NH}_4\text{Cl}$ . Solvent was removed *in vacuo*. The crude was dissolved in DCM and washed with water. The organic layer was collected, dried over anhydrous  $\text{Na}_2\text{SO}_4$ , filtered and the filtrate was concentrated. It was purified by flash column chromatography (11:9 EtOAc:Hex) to yield compound **22** (2.76 g, 53% over two steps).  $R_f = 0.23$  (1:19 MeOH:DCM).  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  9.31 (s, 1H), 8.75 (s, 1H), 8.16 - 7.97 (m, 2H), 7.60 - 7.44 (m, 3H), 6.47 (d,  $J = 2.8$  Hz, 1H), 6.05 - 5.91 (m, 1H),



5.34 (d,  $J = 17.1$  Hz, 1H), 5.27 – 5.20 (m, 1H), 4.60 (d,  $J = 2.2$  Hz, 1H), 4.47 (s, 1H), 2.87 (dd,  $J = 5.6, 2.4$  Hz, 1H), 2.58 – 2.48 (m, 1H), 1.20 (dd,  $J = 21.0, 18.2$  Hz, 3H), 1.12 – 0.98 (m, 18H).  $^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  164.86, 152.66, 151.69, 149.68, 141.81, 135.67, 133.73, 132.80, 128.87, 128.00, 123.86, 118.18, 89.06, 85.04, 76.14, 40.34, 29.36, 18.04, 12.12. HRMS (ESI-TOF) calcd for  $[\text{M} + \text{H}]^+$  508.2744, found: 508.2730.

**Compound 23:** Compound **22** (2.5 g, 4.92 mmol) and NMO (1.47 g, 12.55 mmol) were taken in DCM (125 mL). To this 2% aq. solution of  $\text{OsO}_4$  (0.13 mL, 0.01 mmol) was added. The reaction was allowed to stir for 10 hrs at ambient temperature. The reaction was quenched with satd. aq. solution of  $\text{Na}_2\text{S}_2\text{O}_3$ . DCM was added to the solution and was washed by water. The organic layer was collected, dried over anhydrous  $\text{Na}_2\text{SO}_4$ , filtered and the filtrate was concentrated in vacuo. It was purified by flash column chromatography (1:24 MeOH:DCM) to yield compound **23** (1.79 g, 67%).  $R_f = 0.31$  (1:19 MeOH:DCM). HRMS (ESI-TOF) calcd for  $[\text{M} + \text{H}]^+$  542.28010, found: 542.279322.

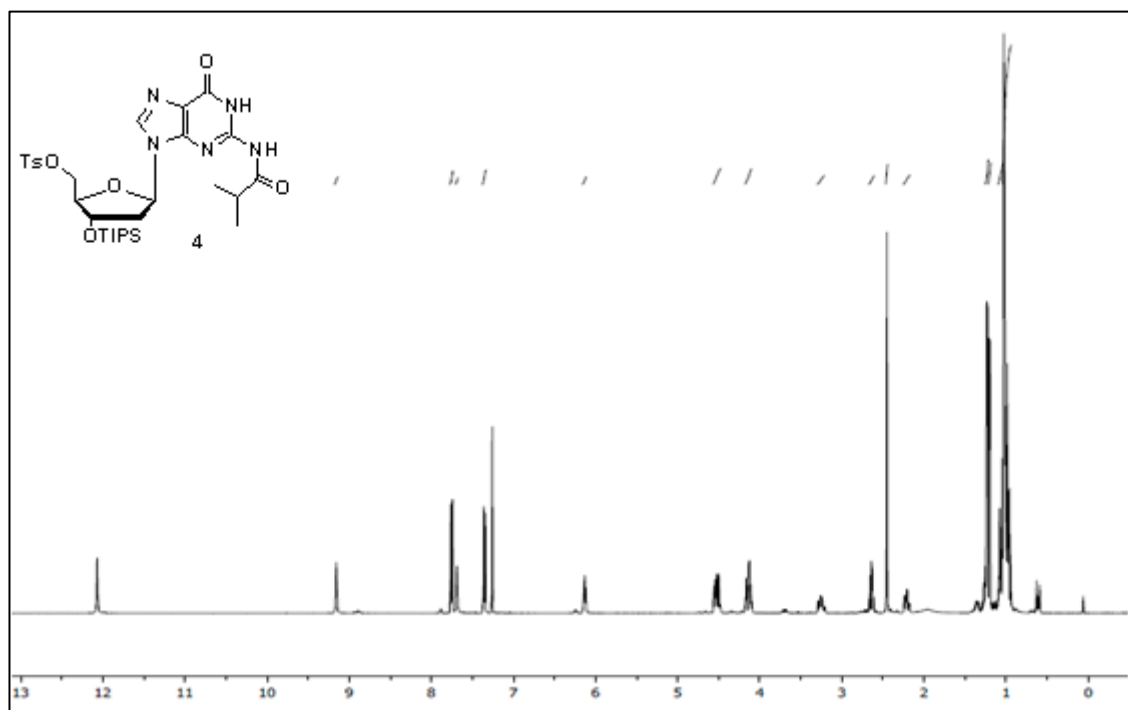
**Compound 24:** Compound **23** (1.70 g, 3.14 mmol) and DMAP (704 mg, 6.28 mmol) were dissolved in DCM (30 mL) and it was cooled to 0 °C. A solution of p-TsCl (659 mg, 3.45 mmol) in DCM (10 mL) was added to the reaction mixture using a dropping funnel. It was slowly allowed to warm to room temperature. The reaction mixture was stirred for 20 hrs at ambient temperature. DCM was added to the solution and was washed by water. The organic layer was collected, dried over anhydrous  $\text{Na}_2\text{SO}_4$ , filtered and the filtrate was concentrated *in vacuo*. It was purified by gravity column chromatography (1.5:98.5 MeOH:DCM) to yield two diastereomers of compound **24**- **24a** (787 mg, 36%) and **24b** (764 mg, 35%).  $R_f$  of **24a** (the higher spot on TLC) = 0.49 (1:19 MeOH:DCM).  $R_f$  of **24b**

(the lower spot on TLC) = 0.47 (1:19 MeOH:DCM). Characterization of compound **24a**:  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.62 (s, 1H), 8.08 (s, 1H), 8.02 (dt,  $J$  = 3.3, 2.4 Hz, 2H), 7.76 – 7.73 (m, 2H), 7.64 – 7.60 (m, 1H), 7.55 – 7.51 (m, 2H), 7.28 (dd,  $J$  = 8.6, 0.6 Hz, 2H), 6.35 – 6.32 (m, 1H), 4.76 (d,  $J$  = 4.7 Hz, 1H), 4.26 (s, 1H), 4.02 (t,  $J$  = 11.6 Hz, 3H), 2.96 (ddd,  $J$  = 13.1, 9.8, 5.0 Hz, 1H), 2.42 (s, 3H), 2.32 – 2.27 (m, 1H), 1.15 – 1.09 (m, 3H), 1.09 – 1.06 (m, 18H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  164.63, 151.92, 150.51, 145.00, 142.63, 133.49, 133.13, 132.94, 129.92, 129.07, 128.05, 124.71, 88.78, 88.20, 74.93, 70.02, 69.89, 41.41, 29.41, 21.77, 18.09, 18.09, 12.05. HRMS (ESI-TOF) calcd for  $[\text{M} + \text{H}]^+$  696.288172, found: 696.28870. Characterization of compound **24b**:  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.71 (s, 1H), 8.08 (s, 1H), 8.04 – 7.99 (m, 2H), 7.82 – 7.78 (m, 2H), 7.64 – 7.59 (m, 1H), 7.55 – 7.50 (m, 2H), 7.36 – 7.32 (m, 2H), 6.32 (dd,  $J$  = 9.6, 5.4 Hz, 1H), 4.77 (d,  $J$  = 4.3 Hz, 1H), 4.19 – 4.09 (m, 4H), 3.07 (ddd,  $J$  = 13.1, 9.7, 4.6 Hz, 1H), 2.44 (s, 3H), 2.31 (dd,  $J$  = 12.9, 5.6 Hz, 1H), 1.13 – 1.03 (m, 21H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  164.60, 152.18, 150.72, 150.41, 145.19, 142.88, 133.59, 133.08, 132.75, 130.03, 129.05, 128.18, 128.02, 124.65, 90.02, 87.46, 72.89, 70.22, 70.17, 41.14, 29.41, 21.79, 18.15, 12.22. HRMS (ESI-TOF) calcd for  $[\text{M} + \text{H}]^+$  696.288172, found: 696.28950.

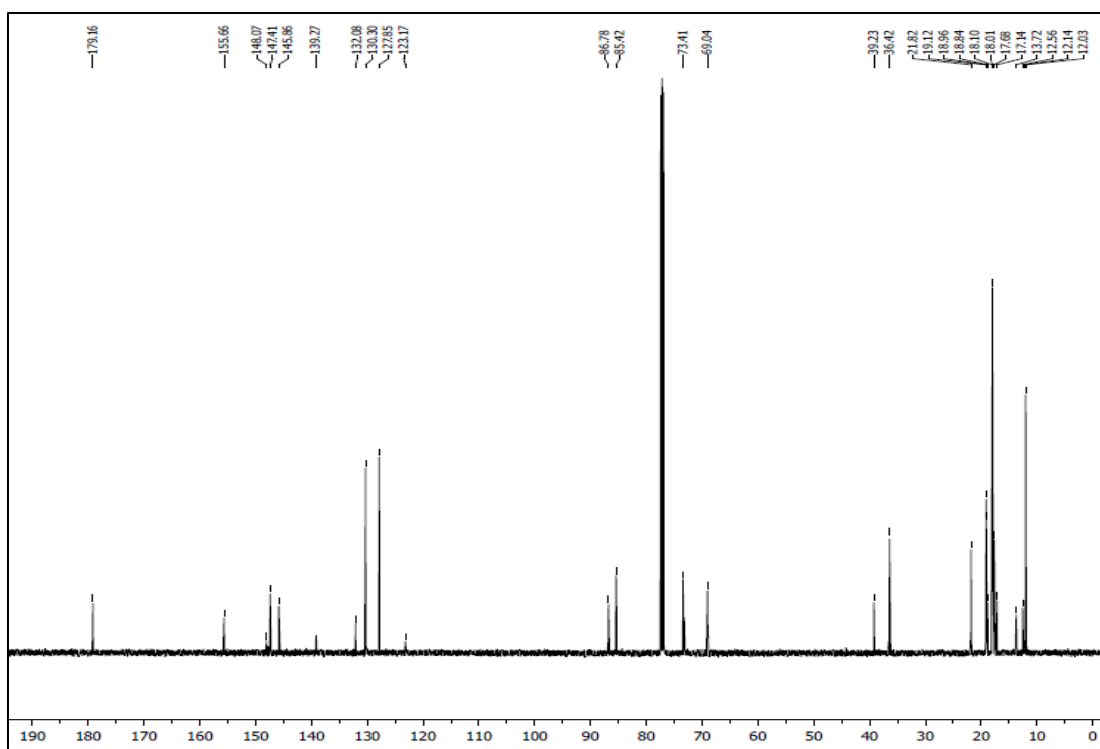
Compound **25**: Both the epimers, **24a** (higher spot) and **24b** (lower spot) were treated with NaI (5 eqvt.) in anhydrous acetone (0.1 M reaction concentration with respect to the starting material) in refluxing condition for 12 hrs. The TLC indicated the complete consumption of the starting material. The reaction mixture was filtered through celite to remove excess NaI. The filtrate was concentrated *in vacuo* and purified by flash column chromatography. Yield **25a** and **25b** was 89% and 86% respectively. Characterization of

compound **25a**:  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.76 – 8.73 (m, 1H), 8.10 (s, 1H), 8.02 (dt,  $J = 8.5, 1.6$  Hz, 2H), 7.64 – 7.59 (m, 1H), 7.53 (dt,  $J = 7.3, 1.7$  Hz, 2H), 6.39 (dd,  $J = 9.8, 5.3$  Hz, 1H), 4.74 (d,  $J = 4.7$  Hz, 1H), 4.65 (s, 1H), 3.98 (d,  $J = 5.1$  Hz, 1H), 3.28 (dd,  $J = 9.7, 5.6$  Hz, 1H), 3.18 (t,  $J = 9.6$  Hz, 1H), 3.05 (ddd,  $J = 13.0, 9.9, 5.0$  Hz, 1H), 2.29 (dd,  $J = 13.0, 5.4$  Hz, 1H), 1.19 – 1.12 (m, 3H), 1.12 – 1.07 (m, 18H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  164.61, 151.94, 150.59, 150.56, 142.78, 133.51, 133.13, 129.07, 128.03, 124.82, 89.17, 88.26, 75.57, 72.86, 41.14, 18.13, 12.10, 6.32. HRMS (ESI-TOF) calcd for  $[\text{M} + \text{H}]^+$  652.1816, found: 652.1819. Characterization of compound **25b**:  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.78 (s, 1H), 8.09 (s, 1H), 8.04 – 8.00 (m, 2H), 7.64 – 7.59 (m, 1H), 7.56 – 7.50 (m, 2H), 6.34 (dd,  $J = 9.7, 5.4$  Hz, 1H), 4.79 (d,  $J = 4.4$  Hz, 1H), 4.27 (d,  $J = 3.1$  Hz, 1H), 4.09 (td,  $J = 8.0, 3.0$  Hz, 1H), 3.29 (qd,  $J = 10.3, 6.4$  Hz, 2H), 3.14 (ddd,  $J = 13.1, 9.8, 4.7$  Hz, 1H), 2.32 (dd,  $J = 13.0, 5.5$  Hz, 1H), 1.14 (ddd,  $J = 10.3, 7.2, 4.0$  Hz, 3H), 1.10 (dd,  $J = 6.8, 3.9$  Hz, 18H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  164.58, 152.30, 150.82, 150.31, 142.90, 133.59, 133.07, 129.06, 128.02, 124.76, 91.26, 87.54, 72.79, 72.67, 41.15, 18.19, 12.29, 6.83. HRMS (ESI-TOF) calcd for  $[\text{M} + \text{H}]^+$  652.1816, found: 652.1841.

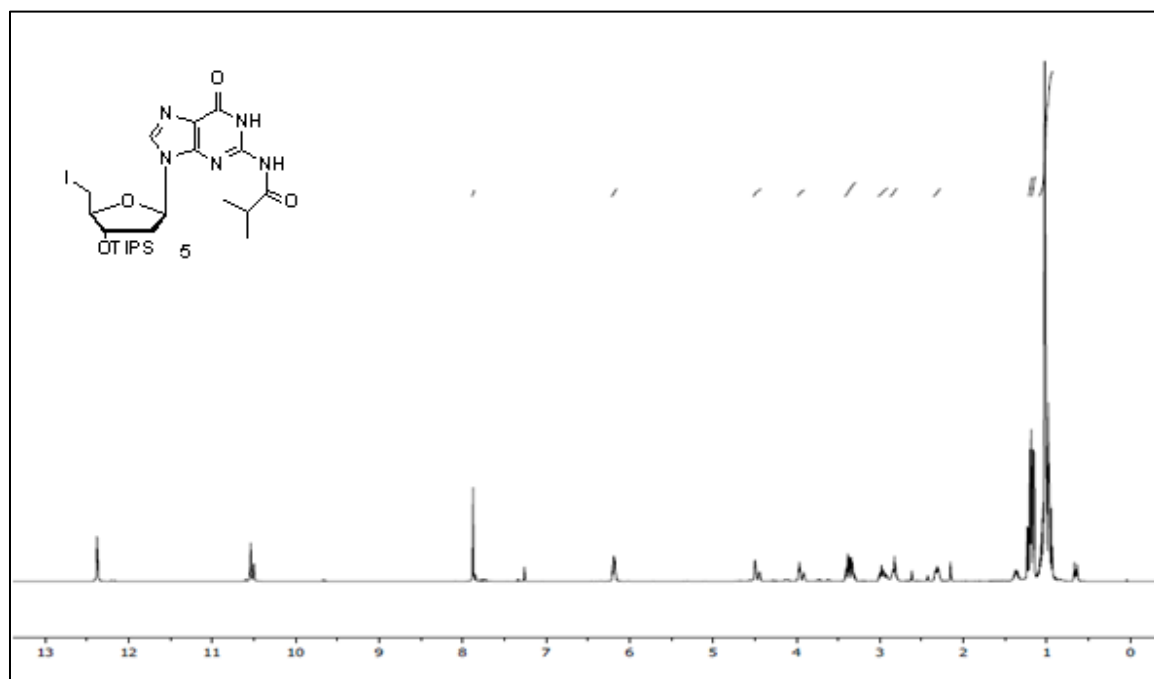
<sup>1</sup>H NMR of Compound 4



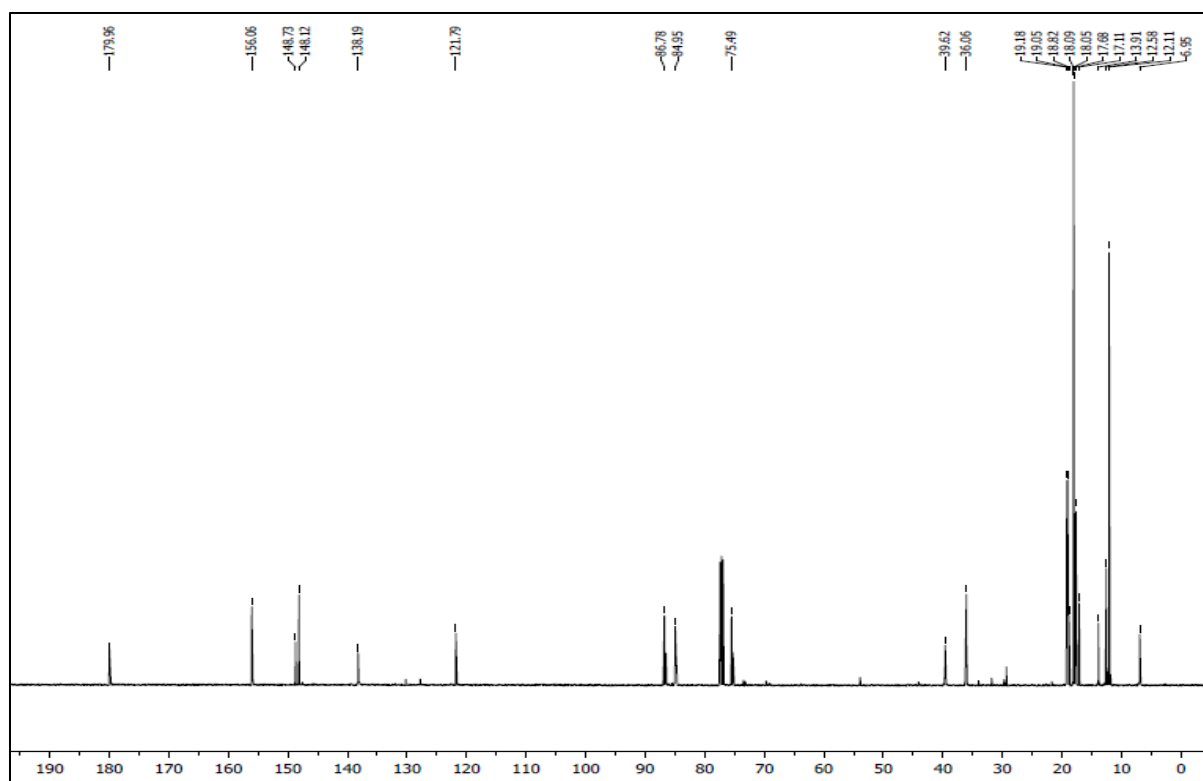
<sup>13</sup>C NMR of Compound 4



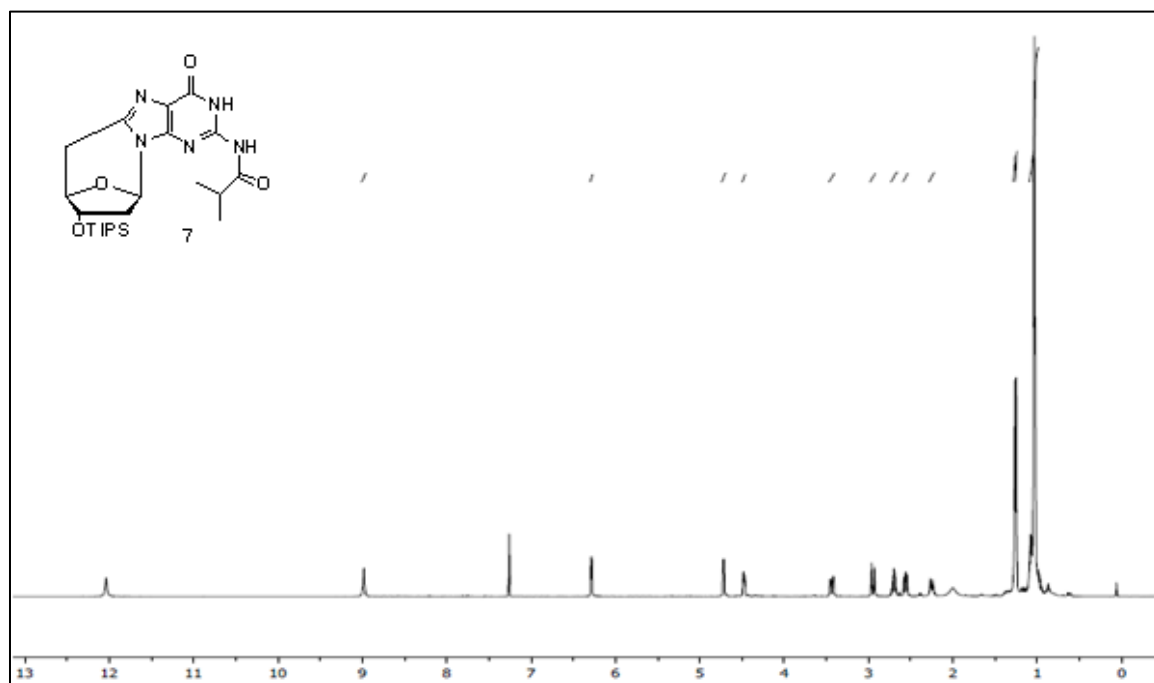
$^1\text{H}$  NMR of Compound **5**



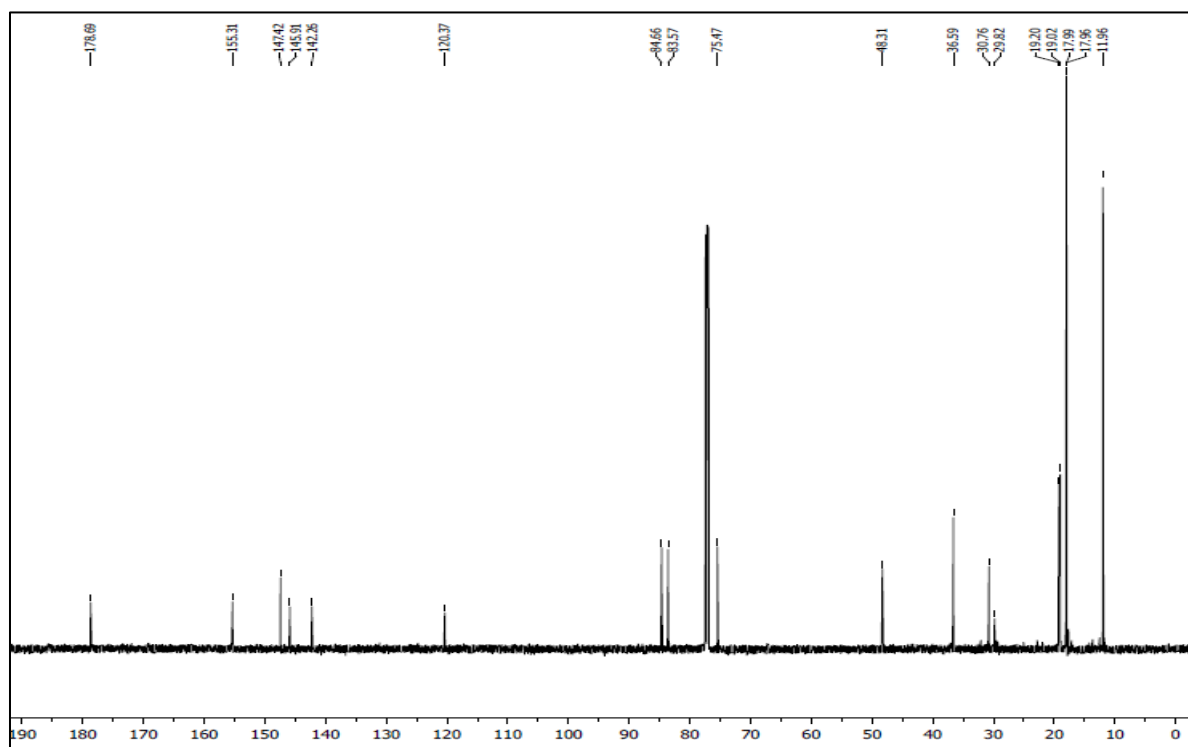
$^{13}\text{C}$  NMR of Compound **5**



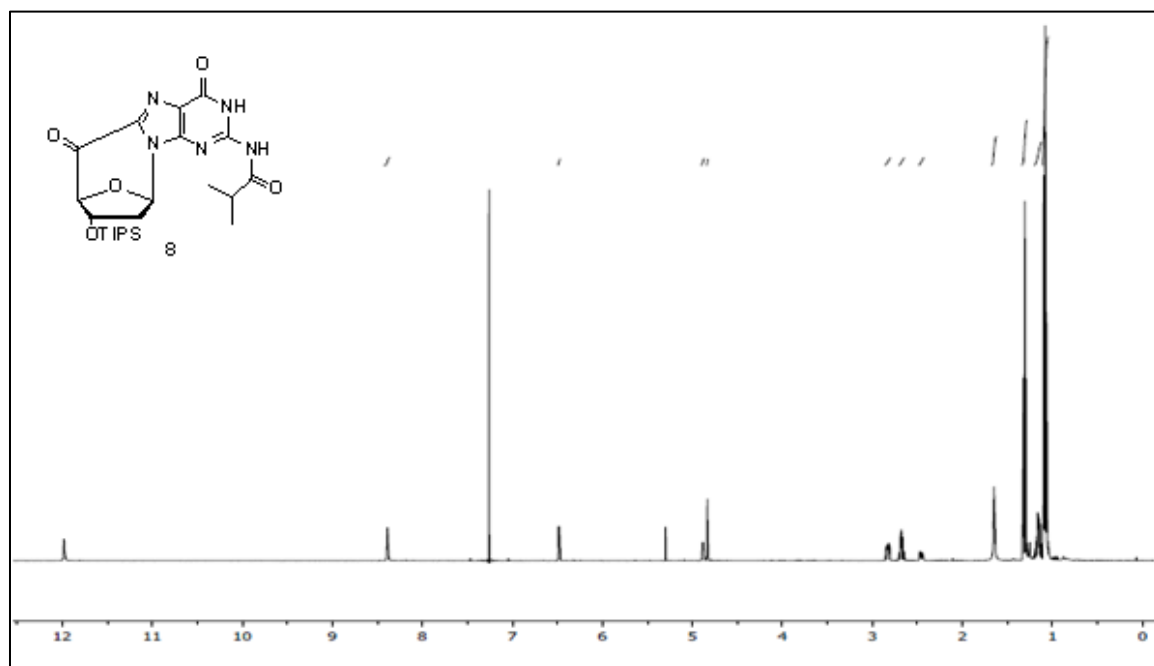
<sup>1</sup>H NMR of Compound 7



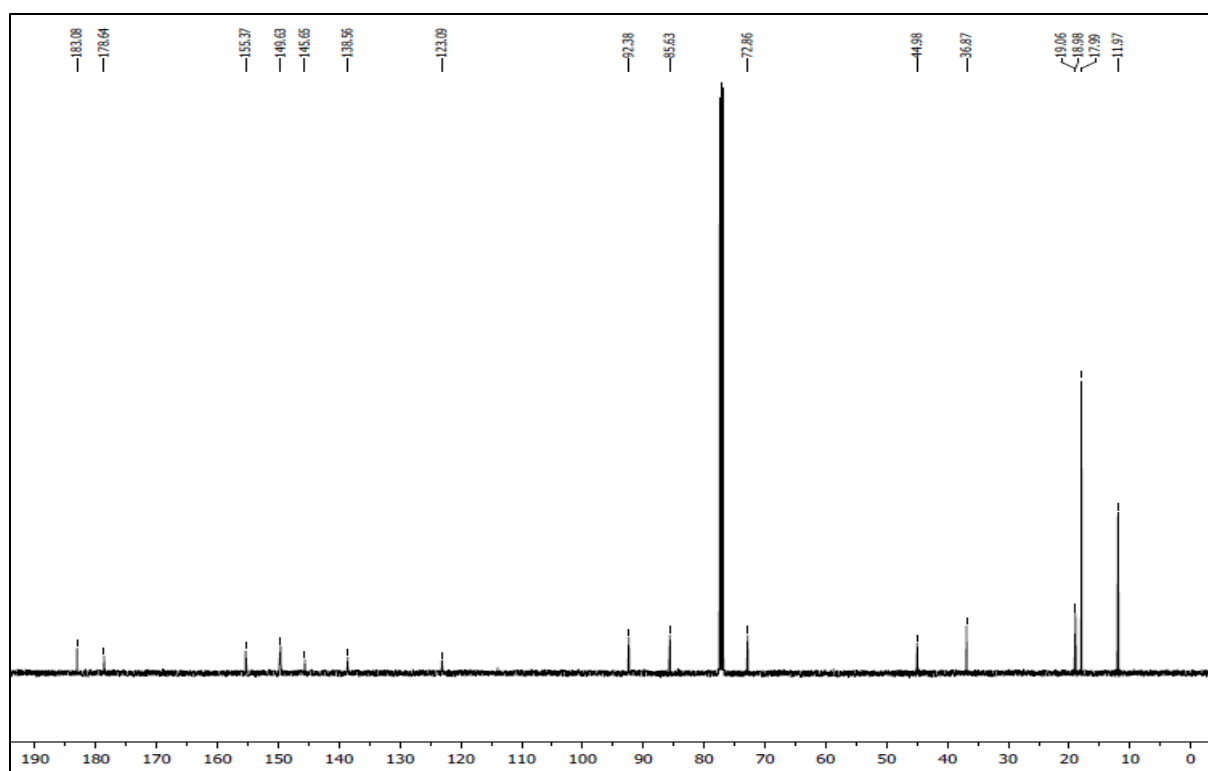
<sup>13</sup>C NMR of Compound 7



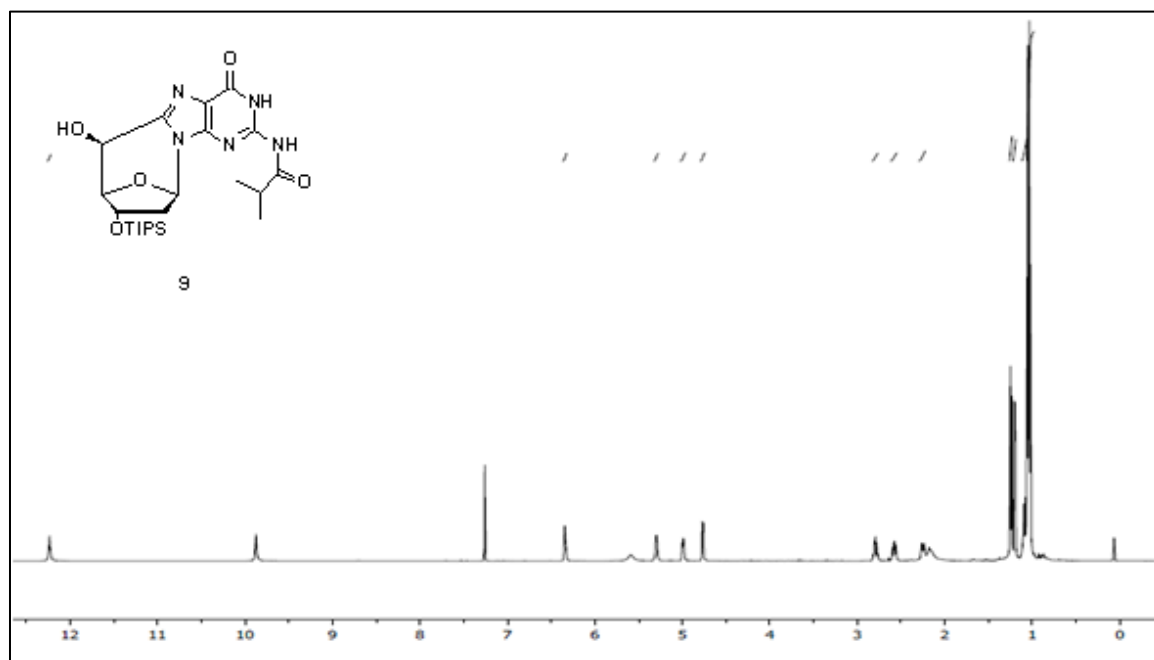
$^1\text{H}$  NMR of Compound **8**



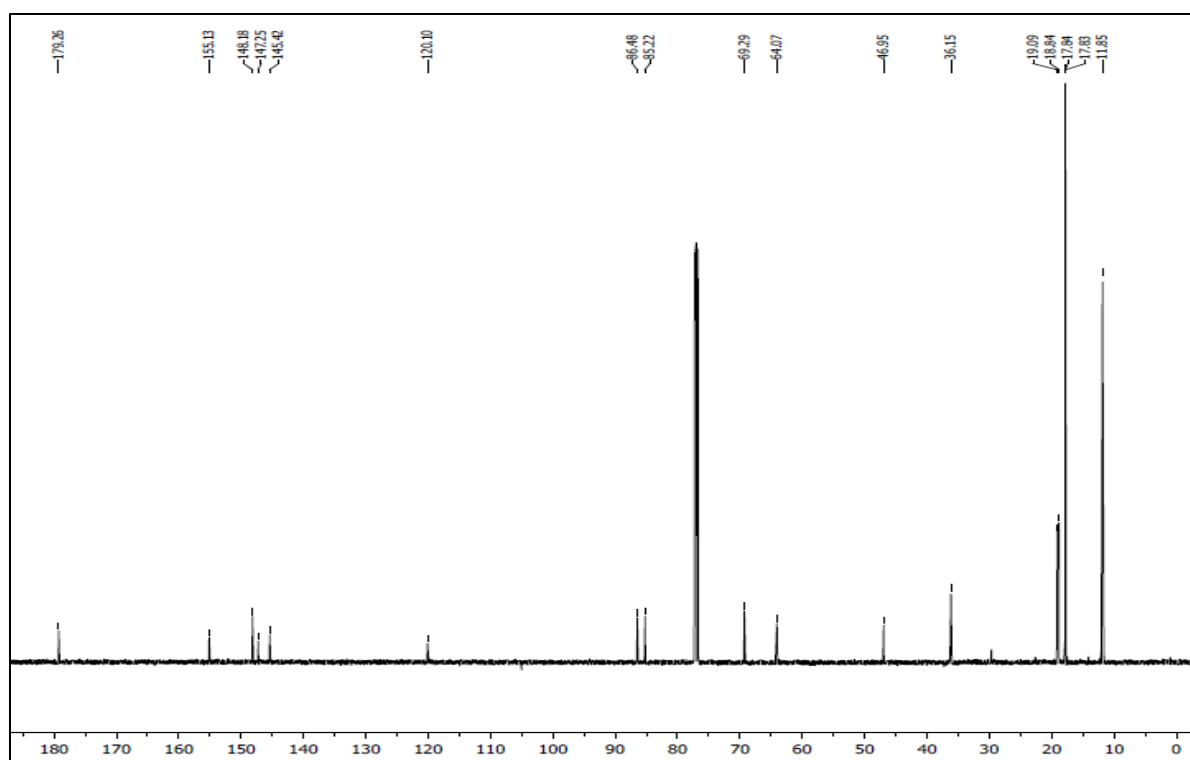
$^{13}\text{C}$  NMR of Compound **8**



$^1\text{H}$  NMR of Compound **9**



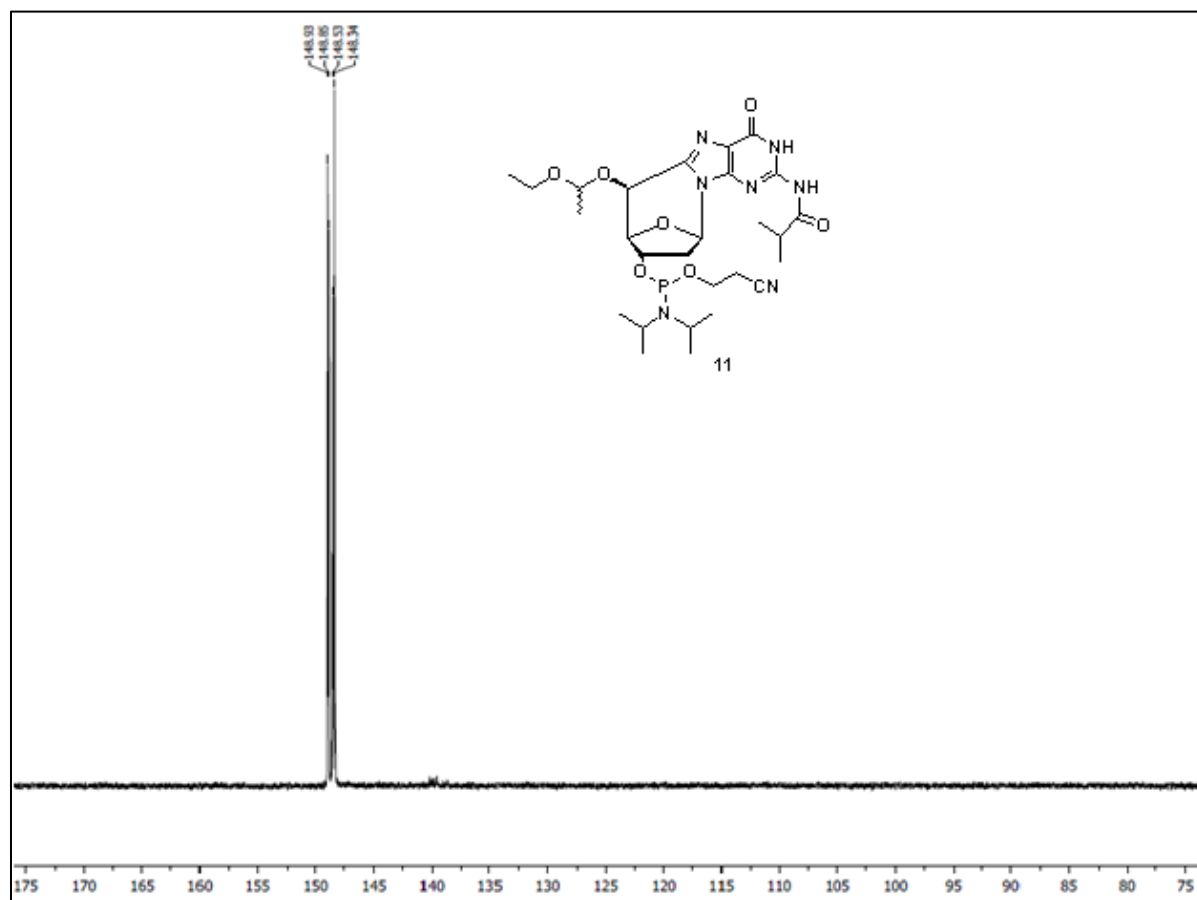
$^{13}\text{C}$  NMR of Compound **9**



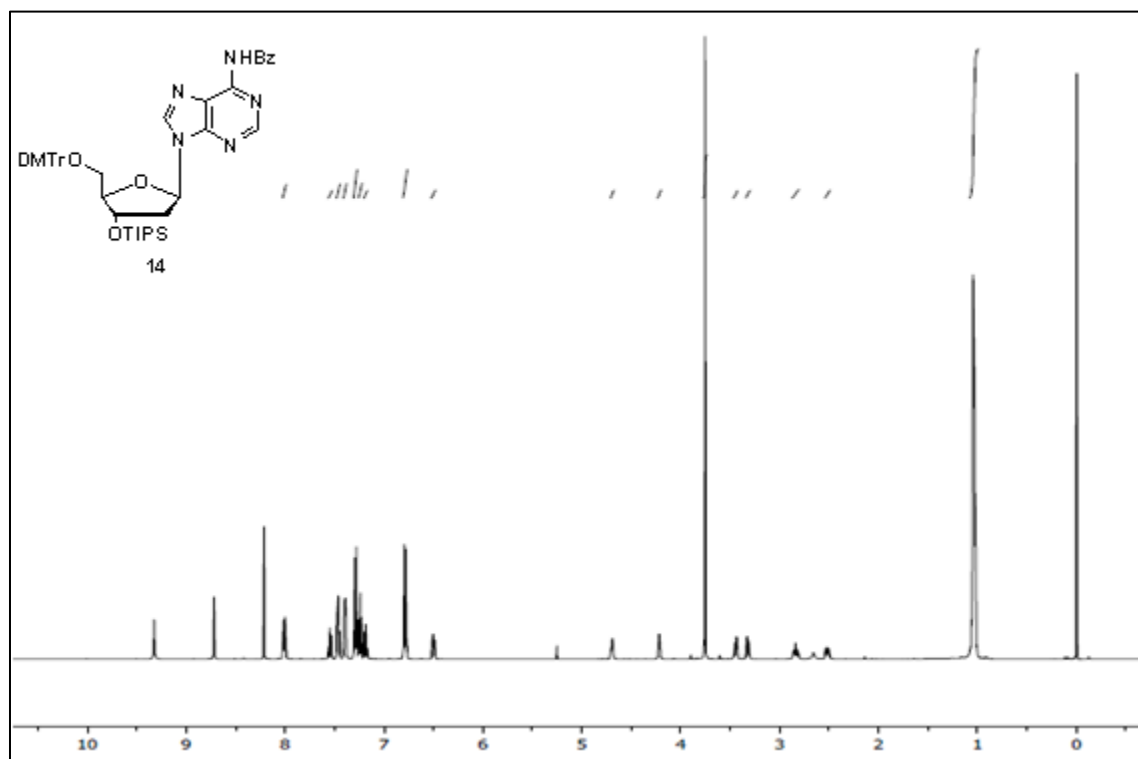




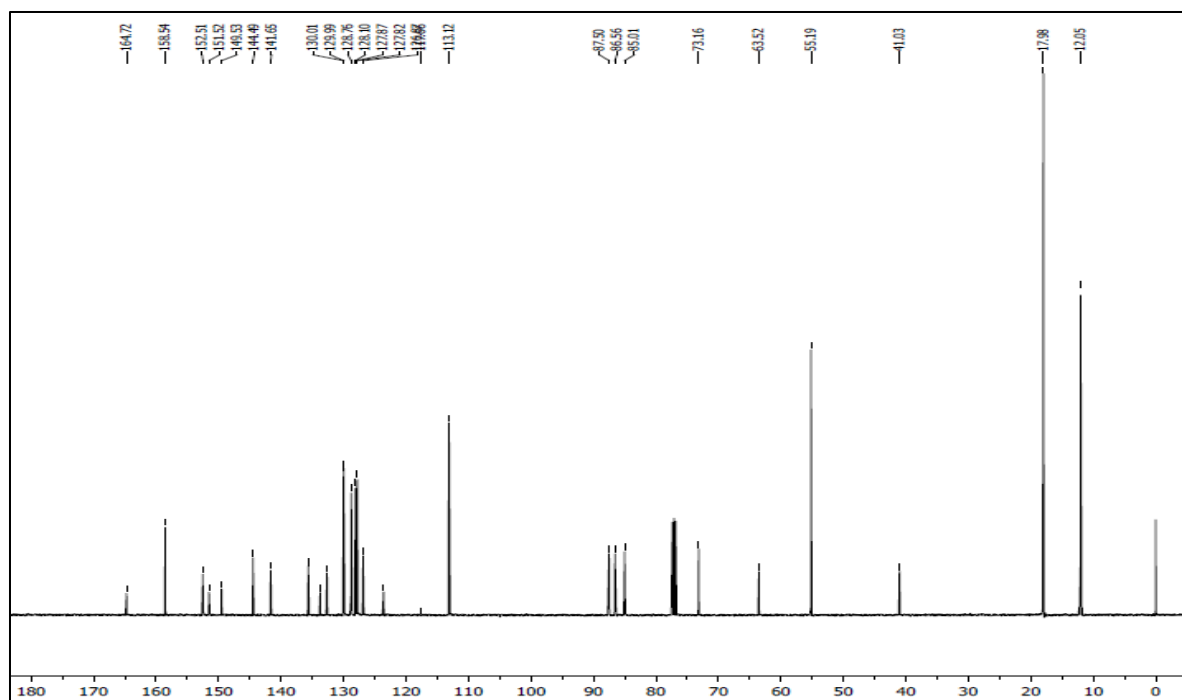
$^{13}\text{P}$  NMR of Compound **11**



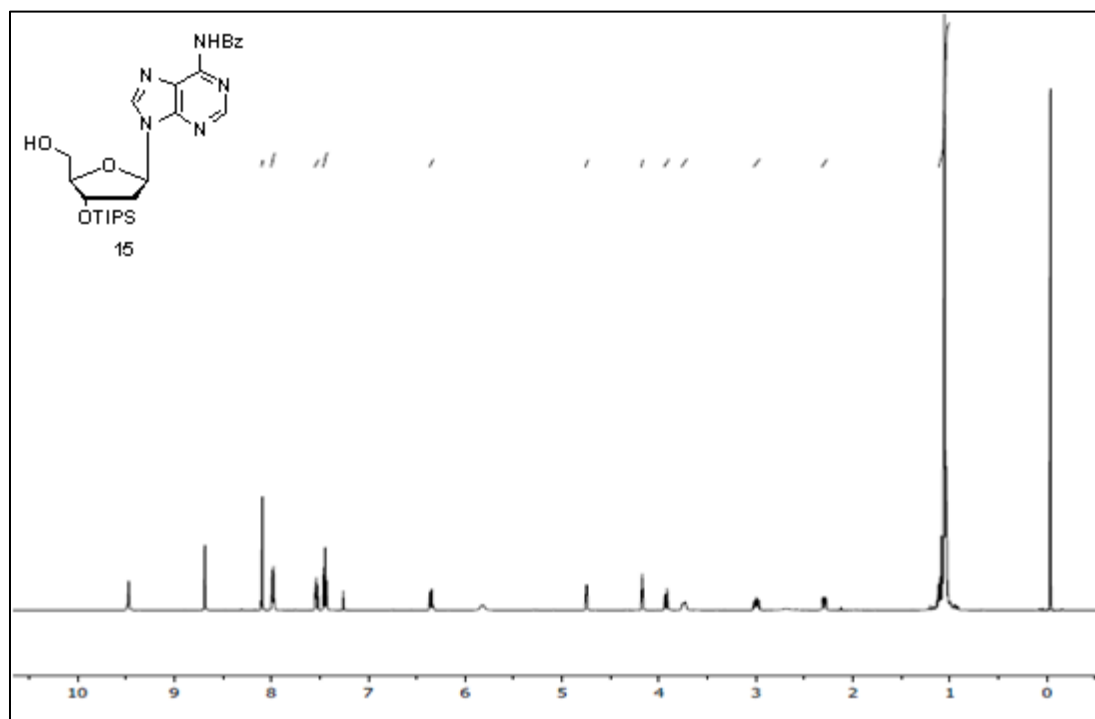
$^1\text{H}$  NMR of Compound **14**



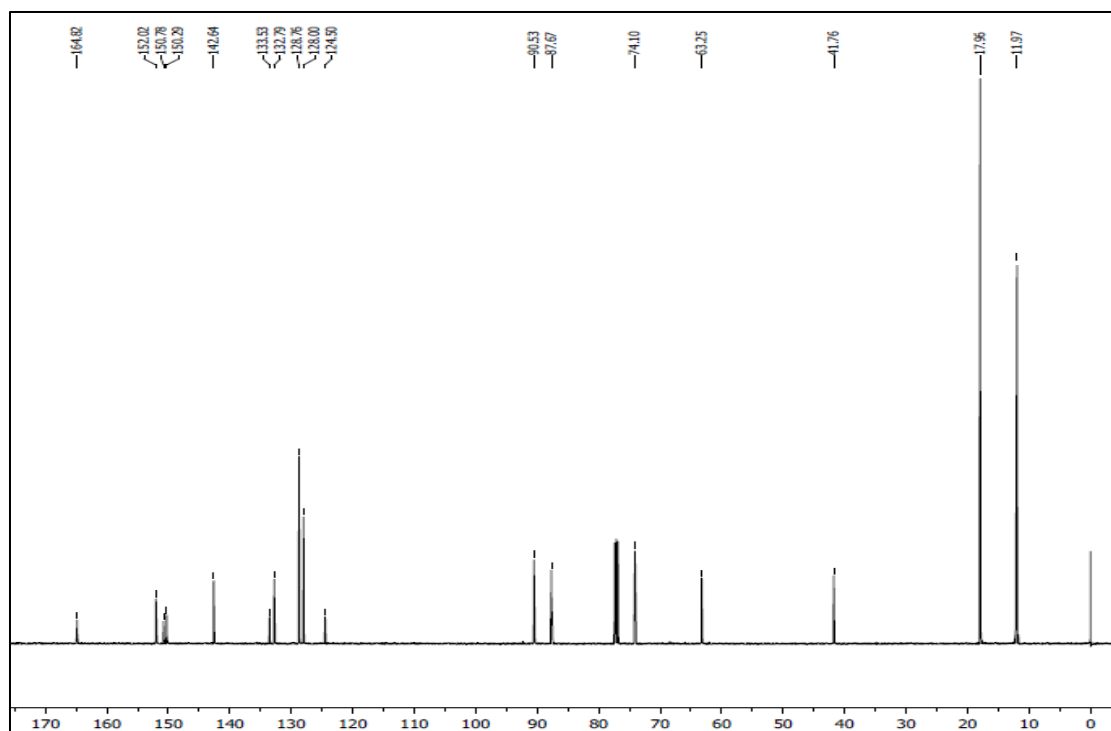
$^{13}\text{C}$  NMR of Compound **14**



<sup>1</sup>H NMR of Compound **15**



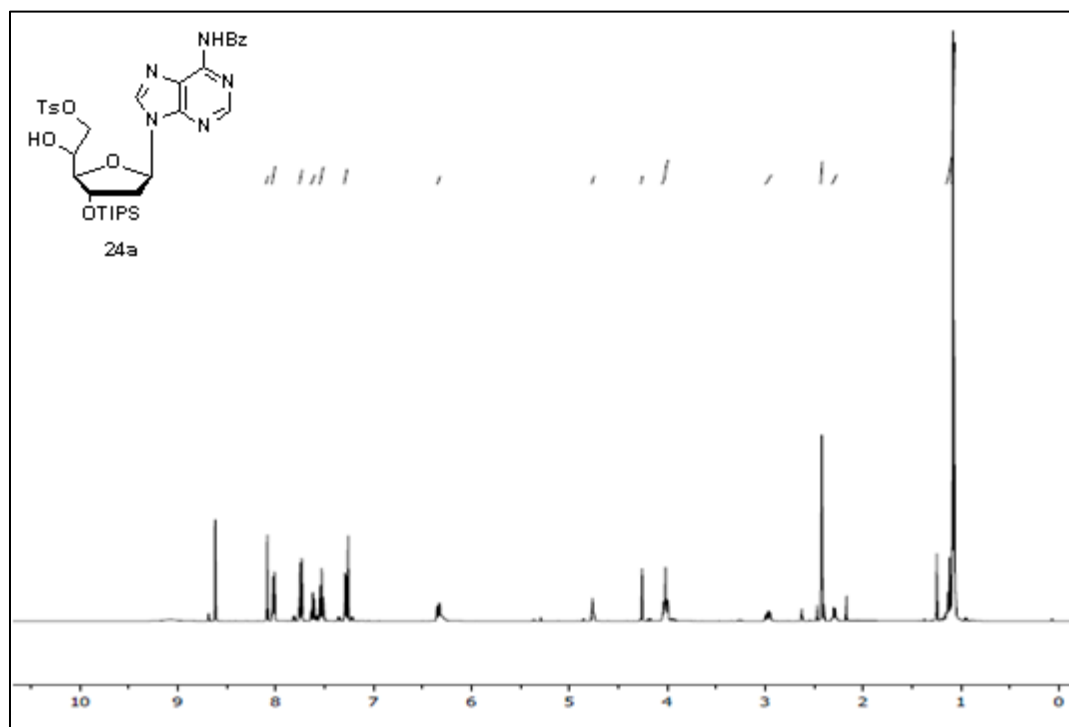
<sup>13</sup>C NMR of Compound **15**



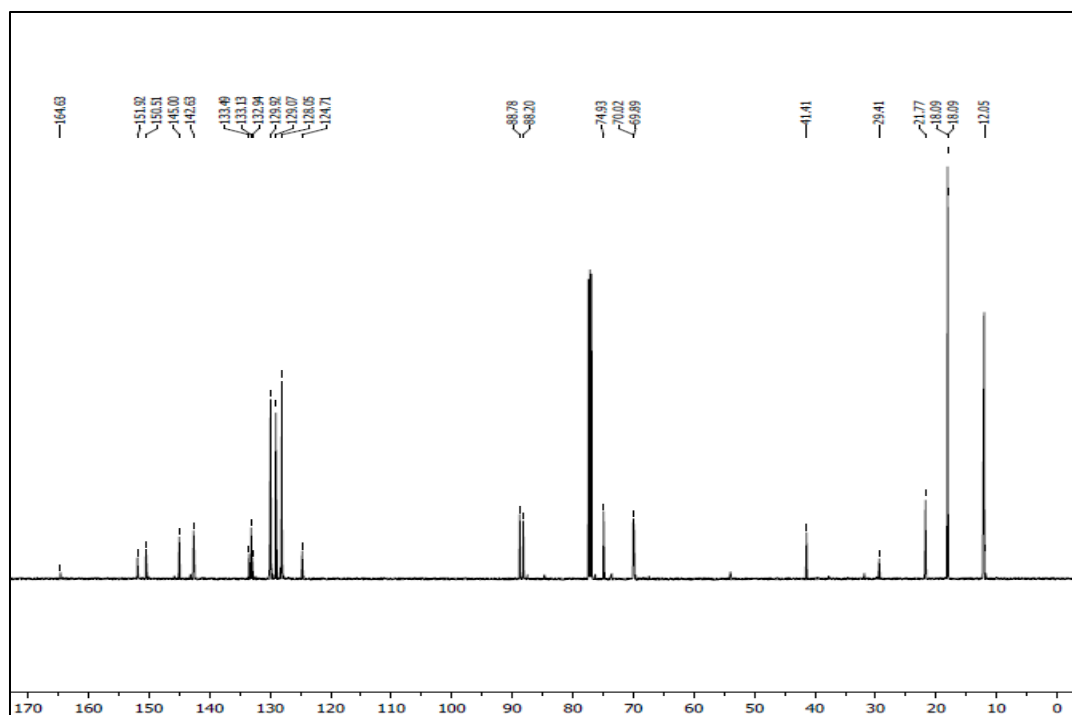
C=CC1OC(C2=CN(C(=N2)N3C=NC=NC3=NBz)N1)OSi(C)(C)C(C)(C)C
  
 22

164.86  
152.66  
151.69  
149.88  
141.81  
135.67  
133.73  
132.80  
128.87  
128.00  
123.86  
118.18  
89.06  
85.04  
76.14  
40.34  
29.36  
19.04  
12.12

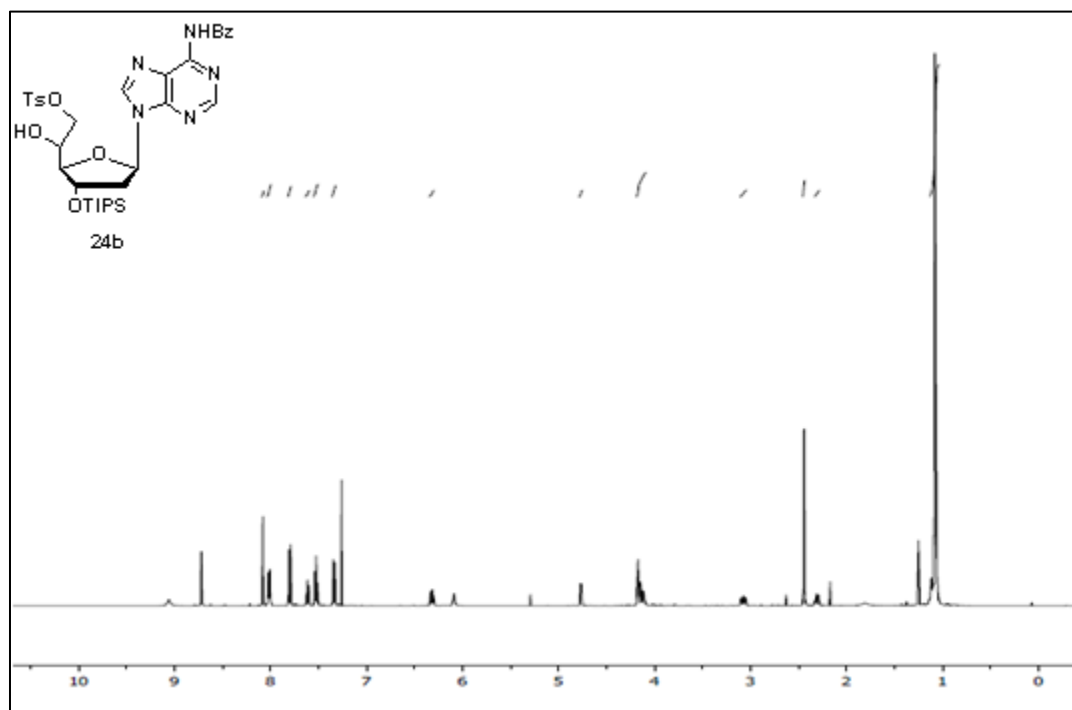
$^1\text{H}$  NMR of Compound **24a**



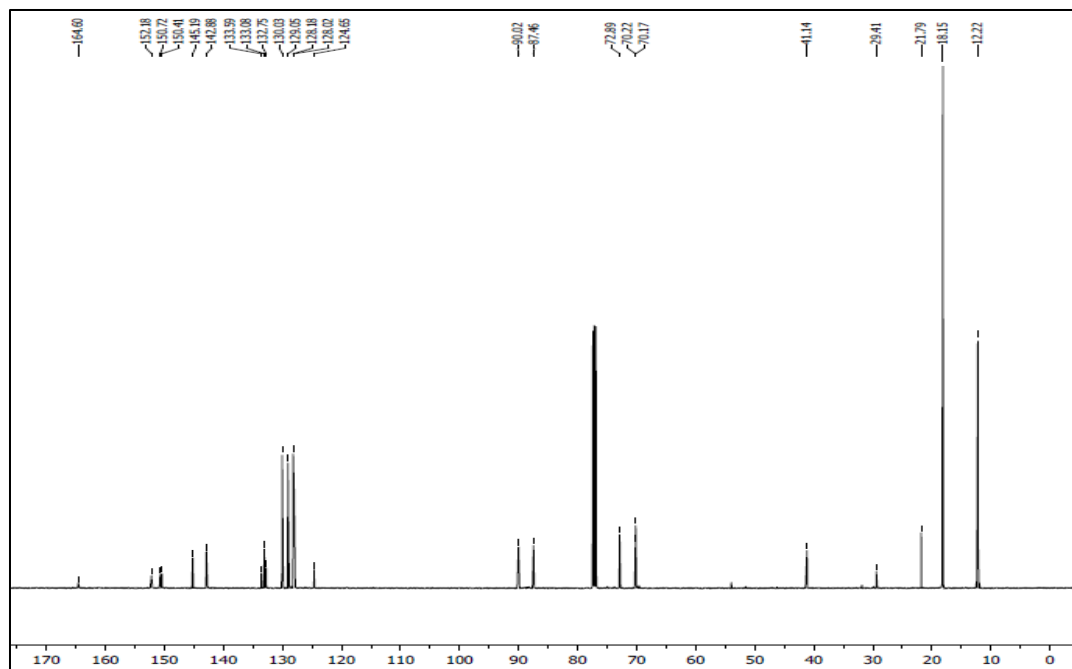
$^{13}\text{C}$  NMR of Compound **24a**



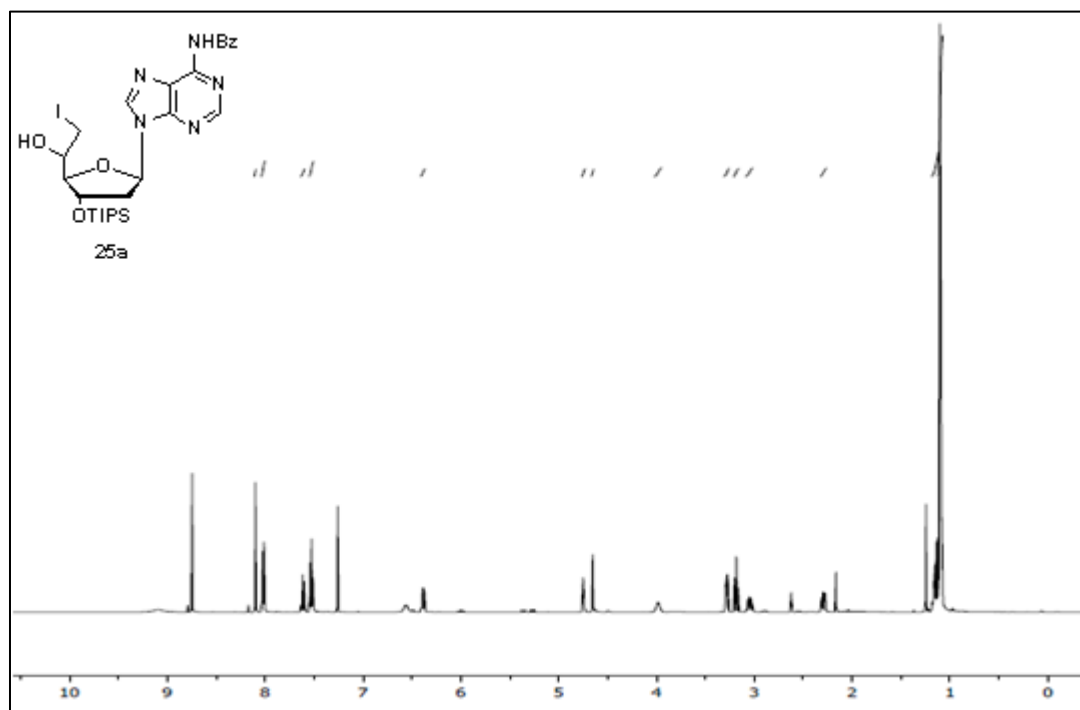
<sup>1</sup>H NMR of Compound **24b**



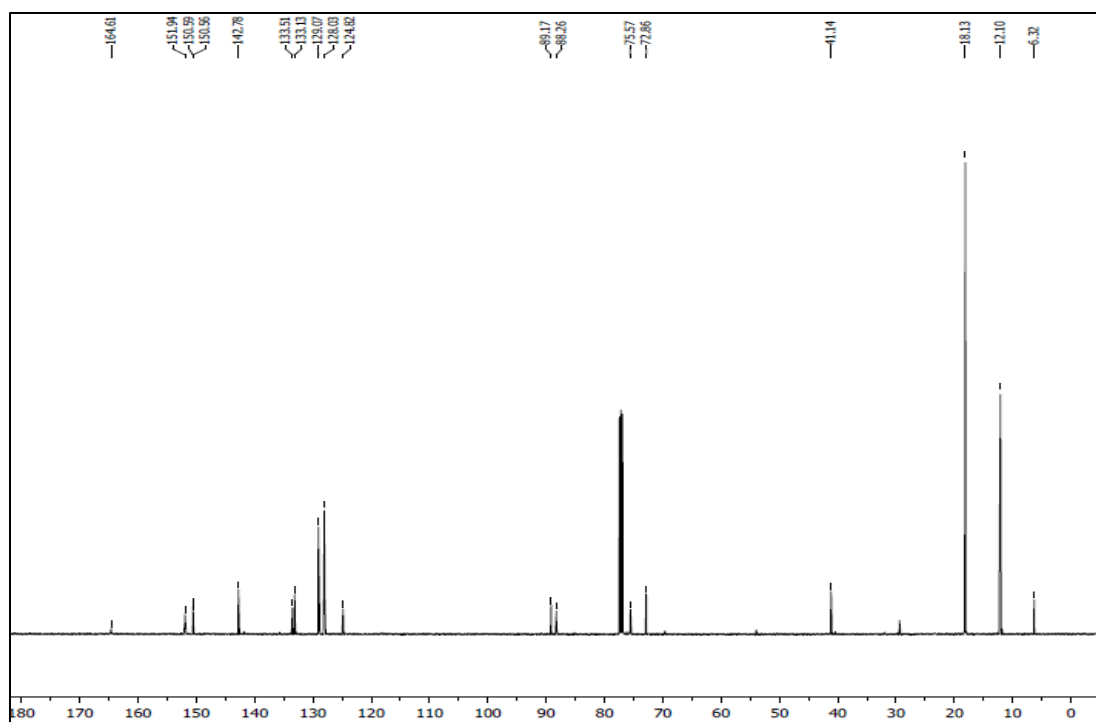
<sup>13</sup>C NMR of Compound **24b**



$^1\text{H}$  NMR of Compound **25a**

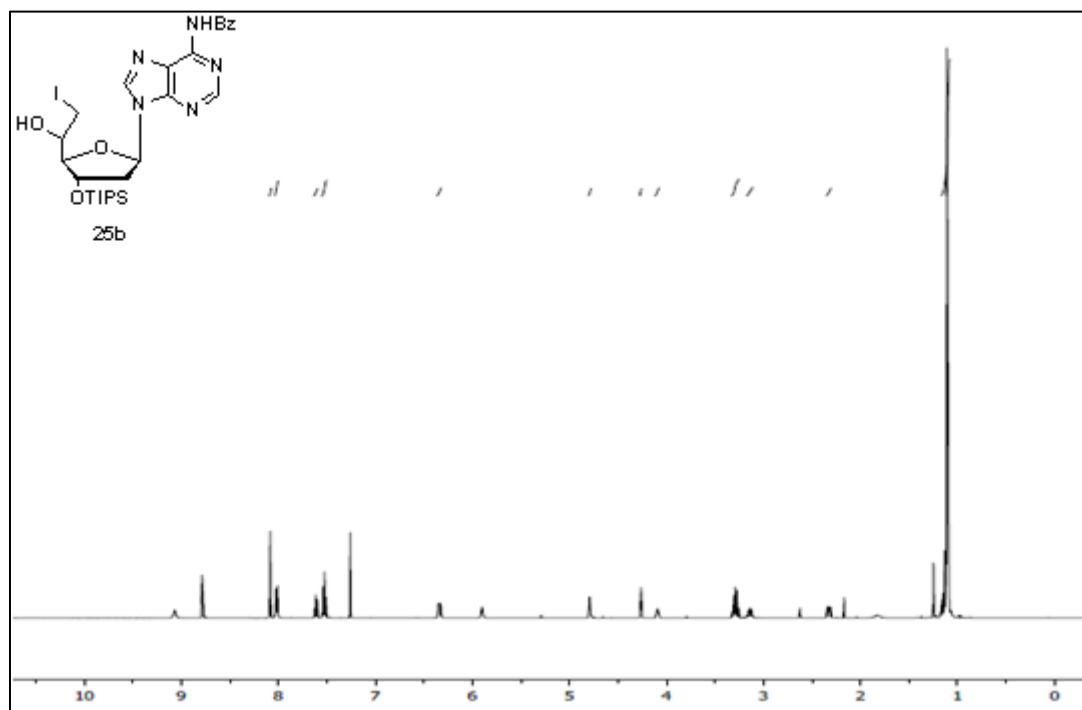


$^{13}\text{C}$  NMR of Compound **25a**

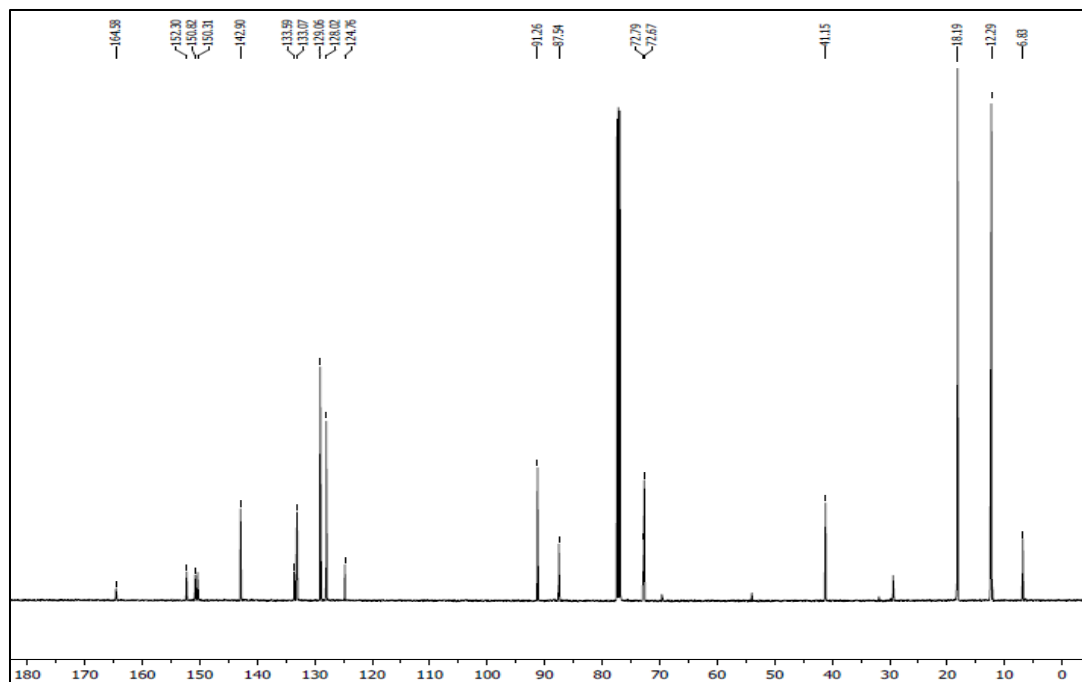




$^1\text{H}$  NMR of Compound **25b**



$^{13}\text{C}$  NMR of Compound **25b**



---

## References

- <sup>1</sup> a) Tarkoy, M.; Bolli, M.; Schweizer, B.; Leumann, C. *Helv. Chim. Acta* **1993**, 76, 481.  
b) Tarkoy, M.; Leumann, C. *Angew. Chem., Int. Ed. Engl* **1993**, 32, 1432. c) Bolli, M.;  
Leumann, C. *Angew. Chem., Int. Ed. Engl* **1995**, 34, 694.
- <sup>2</sup> a) Eschenmoser, A.; Dobler, M. *Helv. Chim. Acta* **1992**, 75, 218. b) Pitsch, S.;  
Wendeborn, S.; Jaun, B.; Eschenmoser, A. *Helv. Chim. Acta* **1993**, 76, 2161. c) Pitsch,  
S.; Eschenmoser, A. *et al. Helv. Chim. Acta* **1995**, 78, 1621. d) Augustyns, K. *et al.*  
*Nucleic Acids Res* **1992**, 20, 4711. e) Van Aerschot, A.; Verheggen, I.; Hendrix, C.;  
Herdewijn, P. *Angew. Chem., Int. Ed. Engl* **1995**, 34, 1338.
- <sup>3</sup> Singh, S. K.; Nielsen, P.; Koshkin, A. A.; Wengel, J. *Chem. Commun.* **1998**, 456.
- <sup>4</sup> a) Rajwanshi, V. K. *et al. J. Chem. Soc., Perkin Trans.* **1999**, 1, 1407 b) Obika, S. *et al.*  
*Tetrahedron* **2002**, 58, 3039. (c) Pedersen, D. S. *et al. Synthesis* **2002**, 802. d) Sørensen,  
M. D. *et al. J. Am. Chem. Soc.* **2002**, 124, 2164. e) Pedersen, D. S. *et al. Synthesis* **2004**,  
578. f) Christensen, N K. *et al. Chem. Commun.* **2004**, 282. g) Wengel, J. *et al. Org.*  
*Biomol. Chem.* **2004**, 2, 80. h) Gaubert, G. *et al. Tetrahedron* **2006**, 62, 2278. i) Kumar,  
T. S. *et al. J. Org. Chem.* **2006**, 71, 4188.
- <sup>5</sup> a) Baker, B. F.; Lot, S. S.; Condon, T. P.; Cheng-Flournoy, S.; Lesnik, E. A.; Sasmor,  
H. M.; Bennett, C. F. *J. Biol. Chem.* **1997**, 272, 11994. b) Wengel, J. *Acc. Chem. Res.*  
**1999**, 32, 301. c) Obika, S.; Hemamayi, R.; Masuda, T.; Sugimoto, T.; Nakagawa, S.;  
Mayumi, T.; Imanishi, T. *Nucleic Acids Res. Suppl.* **2001**, 1, 145.
- <sup>6</sup> Frieden, M.; Christensen, S. M.; Mikkelsen, N. D.; Rosenbohm, C.; Thruø, C. A.;  
Westergaard, M.; Hansen, H. F.; Orum, H.; Koch, T. *Nucleic Acids Res.* **2003**, 31, 6365.

- 
- <sup>7</sup> Obika, S.; Nanbu, D.; Hari, Y.; Andoh, J.; Morio, K.; Doi, T.; Imanishi, T. *Tetrahedron Lett.* **1998**, *39*, 5401.
- <sup>8</sup> Fluiter, K.; ten Asbroek, A. L.; De Wissel, M. B.; Jakobs, M. E.; Wissenbach, M.; Olsson, H.; Olsen, O.; Oerum, H.; Baas, F. *Nucleic Acids Res.* **2003**, *31*, 953.
- <sup>9</sup> a) Chatgililoglu, C.; Bazzanini, R.; Jimenez, L. B.; Miranda, M. A. *Chem. Res. Toxicol.* **2007**, *20*, 1820–1824. b) Belmadoui, N; Boussicault, F.; Guerra, M.; Ravanat, J. -L.; Chatgililoglu, C.; Cadet, *J.Org. Biomol. Chem.* **2010**, *8*, 3211. c) Romieu, A.; Gasparutto, D.; Molko, D.; Cadet, J. *Nucleosides & Nucleotides* **1999**, *18*(6&7), 1331-1333. d) Romieu, A.; Gasparutto, D.; Molko, D.; Cadet, J. *J. Org. Chem.* **1998**, *63*, 5245-5249.
- <sup>10</sup> Pullman, B.; Pullman, A. *Nature* **1961**, *189*, 725–727.
- <sup>11</sup> Matsuda, A.; Ueda, T *Chem. Pharm. Bull.* **1986**, *34*(4), 1573.
- <sup>12</sup> Omura, K.; Sharma, A. K.; Swern, D. *J. Org. Chem.* **1976**, *41*, 957-962.
- <sup>13</sup> Griffith, W. P. Ley, S. V.; Whitcombe, G. P.; White, A. D. *J. Am. Chem. Soc.* **1987**, *109*, 1625-1627.
- <sup>14</sup> Corey, E. J.; Suggs, J. W. *Tetrahedron Lett.* **1975**, 2647-2650.
- <sup>15</sup> Dess, D. B.; Martin, J. C. *J.Org. Chem.* **1983**, *48*, 4155-4156.
- <sup>16</sup> VanRheenen, V.; Kelly, R. C.; Cha, D. Y. *Tetrahedron Lett.* **1976**, 1973-1976.

## **Chapter 4**

# **Purine-like Janus Wedge Residues**

## Background

Over the last two decades, there has been a huge interest in the field of nucleic acid based technologies for genome analyses, diagnostics and therapeutics. It gained even more momentum after the projection of the Human Genome Project (HGP) in 1990 by the U.S. Department of Energy and the National Institutes of Health. During the early years of the HGP, the Wellcome Trust (U.K.) became a major partner; additional contributions came from Japan, France, Germany, China, and others. Though the HGP is finished, analyses of the data will continue for many years<sup>1</sup>. The completion of this project will effect a revolution in biomedical sciences and related industries. In particular, tailor-made medications, molecular targeted therapies, and genetic medicines based on human genomic information are attracting much attention worldwide<sup>2</sup>. Among them, antigene strategy is one promising technology to regulate gene expression in living cells. There are issues, limitations, roadblocks but, still, significant progress has been made. Hopefully, this technology will be the answer to a number of disease-curing and biomedical problems.

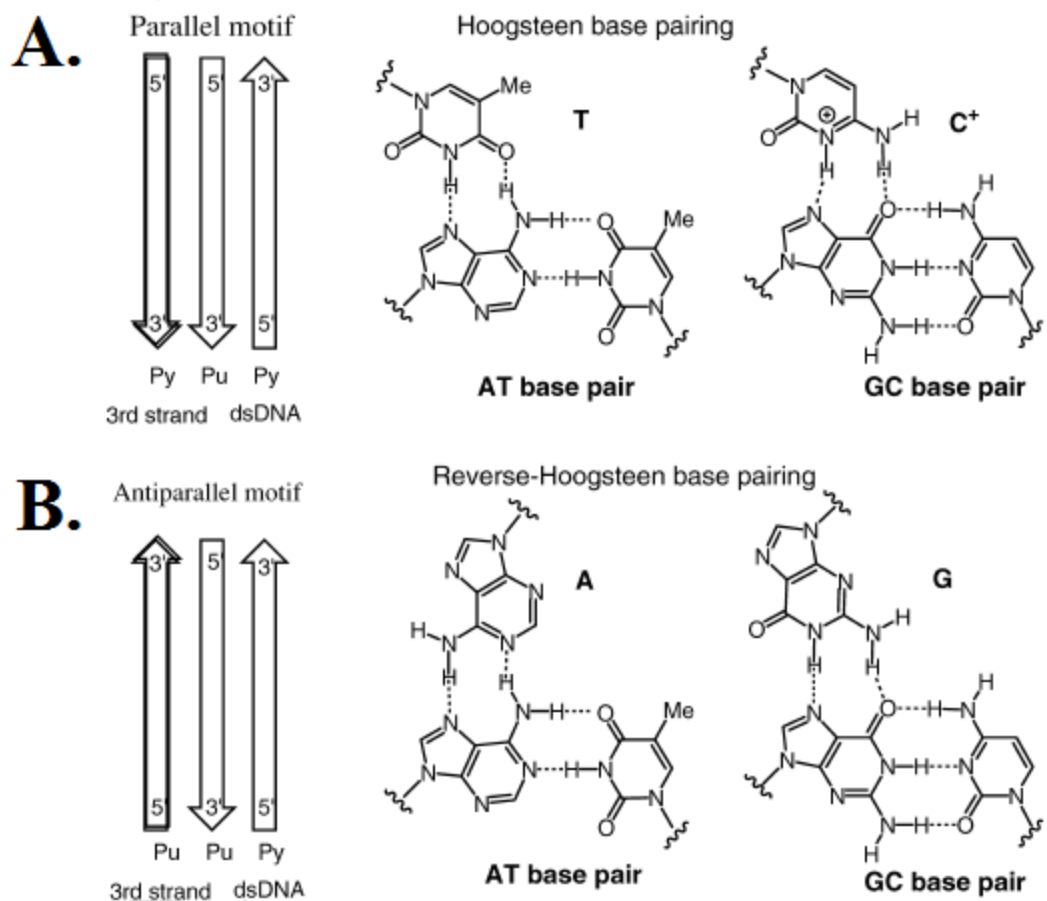
Conceptually, there are two major ways an agent can bind to a DNA duplex: from outside and by invading the DNA duplex. Agents like triplex forming oligonucleotides (TFOs) bind externally in the major groove of duplex DNA with high specificity and affinity<sup>3</sup>. On the other hand, peptide nucleic acids (PNA) are known to invade the DNA duplex and form triplexes; thereby interfere with transcription internally<sup>4</sup>. In the next section we will discuss various aspects of TFOs. The goal of my third project is to target DNA duplex sequence specifically using PNA employing a fundamentally new approach, the Janus Wedge format.

## 4.1. Triplex Forming Oligonucleotides

### 4.1.1. *Types of triplexes*

The fact that nucleic acid can form a triplex structure was observed in 1957<sup>5</sup> and the mode of nucleobase triads was experimentally determined by Hoogsteen<sup>6</sup>. Regardless of the specificity that can be achieved, triplex formation is restricted to DNA duplexes with poly-purine on one strand and pyrimidines on the other. Substantial effort has been expended in the study of the triplex binding code. Though it was initially observed that only the purine oligonucleotides could bind their target under physiologic conditions (most notably neutral pH) the advancement of base substitution technology has allowed the engineering of pyrimidine TFOs which can also bind effectively under physiologic conditions<sup>7</sup>.

In depth research revealed that the triplexes can be divided into two types: parallel and antiparallel<sup>8</sup>. In the parallel motif triplex, the third strand is a homopyrimidine sequence. This third strand and the Pu strand of DNA duplex are in a parallel orientation. The base triads form Hoogsteen hydrogen bondings in the parallel motif triplex (Figure 4.1.A.). On the other hand, the antiparallel-type triplex has the third strand with a homopurine sequence. This third strand and the Py strand of the duplex are in an antiparallel orientation. The base triads have reverse-Hoogsteen hydrogen bondings (Figure 4.1.B.). Of the two triplex motifs, most of the focus was on the parallel motif, in which tuning the properties of the external strand by an appropriate chemical modification is less complicated<sup>9</sup>.



**Figure 4.1.** Schematic illustration of- A. parallel motif triplex and structure of Hoogsteen Base Triads B. anti-parallel motif triplex and structure of Reverse-Hoogsteen Base Triads.

#### 4.1.2. Limitations of TFOs

Although there has been significant progress in using TFOs for regulating transcription, there are a few limitations that appear to be major roadblock for its application as therapeutics. Some of these limitations are mentioned below:

- To form a stable triplex, a homopurine–homopyrimidine tract is required in the target duplex. It also signifies that mixed sequences cannot be targeted. The

occurrence of such unique stretch of polypurines is rare. Hence its applicability is limited.

- Both Hoogsteen and reverse-Hoogsteen triads have only two hydrogen bondings, while the Watson–Crick base pairs have two or three hydrogen bondings. This results in low stability of the triplex structure.
- The cytosine in a parallel motif TFO must be protonated at its N3 position to form a stable nucleobase triad  $C^+*GC$  (in which \* means Hoogsteen bonds). This signifies that the stability of a parallel motif triplex drastically decreases under neutral or basic conditions, although it is reasonably stable under acidic conditions.
- Physiologic potassium concentrations can inhibit binding in particularly G-rich TFOs. Under this condition, G quartets and other secondary structures within the oligonucleotides are favored<sup>10</sup>. This has been addressed with a variety of chemical substitutions. Replacement of natural bases with modified bases such as 6-thioguanine<sup>11</sup> or 7-deazaxanthine<sup>12</sup> has addressed the problem to some extent.

In the next section we will discuss about a relatively new approach of triplex formation using Janus Wedge triplexes.

#### **4.2. Janus Wedge (Jw) Triplexes**

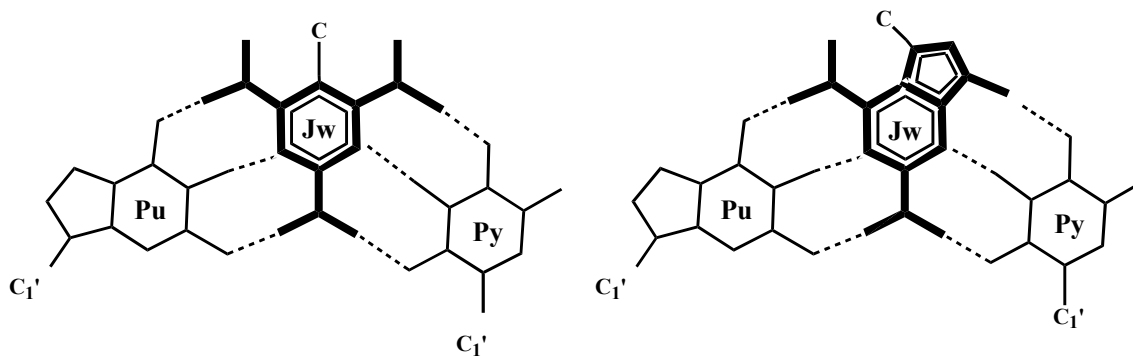
The most of the success with PNA mediated DNA duplex binding relied primarily on the triplex invasion mechanism (Figure 1.20). One of the PNA strands binds with one of the DNA strands of the duplex from the Watson-Crick face and the other PNA strand binds the same DNA strand from the Hoogsteen face (the two PNA strands may be a part of



bis-PNA molecule). As in the case of TFO, a major concern is to have a polypurine stretch in the DNA strand and it mostly fails with a mixed sequence.

This led us to think about alternative approaches. The challenge was can we actually invade the DNA duplex and form a DNA-PNA-DNA triplex. If we can prepare the PNA strand by tethering unnatural bases with proper hydrogen bonding forming profile that can actually differentiate A-T, T-A, G-C and C-G, then we can potentially achieve our goal to target any part of the DNA duplex. The success of this new approach will provide a one-residue to one-base-pair targeting format for the recognition of all four DNA base pairs, and thus provide a motif for generalizing the process of DNA duplex recognition using the triple helix format.

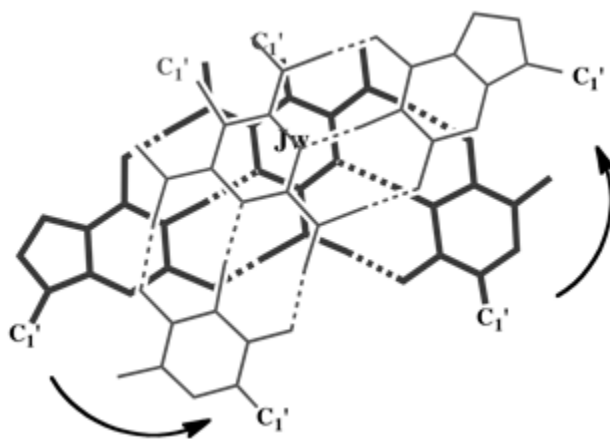
In our approach of Jw triplexes, a third strand of DNA (or PNA) inserts itself between the two Watson-Crick faces of the base pairs of the duplex to result in a new type of DNA triplex. These triplexes are termed Janus Wedge triplexes<sup>13</sup> after the Roman god Janus. Janus stood for the beginning of everything, guardian of heaven and of all gates and doors. Janus is typically pictured with two faces, supposedly so he could see what was coming and going. The Jw residue in the Janus Wedge triplex, contains two hydrogen bonding faces such that after insertion into the DNA duplex, the Jw residue is hydrogen-bonded to both Watson-Crick faces of the target base pair. Both pyrimidine-like (Figure 4.2, left) and purine-like (Figure 4.2, right) residues can be used in this design.



**Figure 4.2.** Schematic representation of pyrimidine (left) and purine (right) like Jw bases.

The Jw triplexes are envisioned as thermodynamically favorable over the DNA duplex because:

- Three-stranded complexes contain significantly more interstrand hydrogen bonds than the DNA duplexes. So the triplex structure should be stable.
- Enhanced base stacking between the consecutive triplex layers, because of the isomorphic structures, should also contribute to the stability of the triplex structure (Figure 4.3).

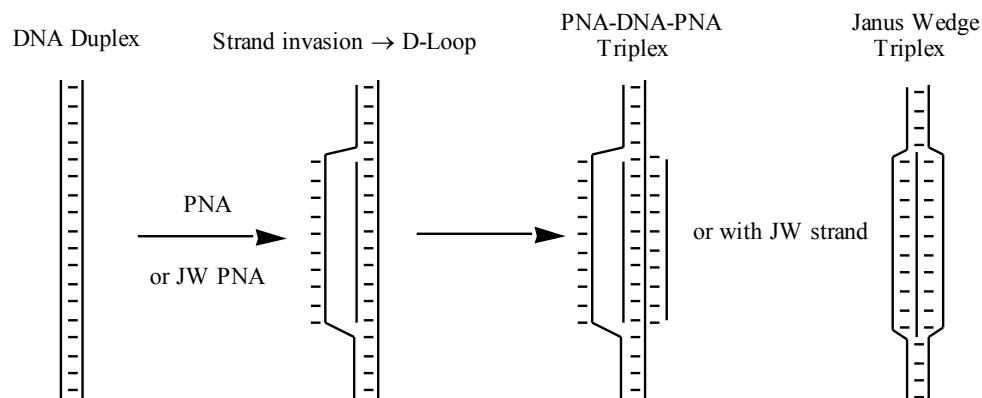


**Figure 4.3.** Enhanced base stacking between the triplex layers.

To investigate the potentiality of this approach, we need two components to prepare these agents:

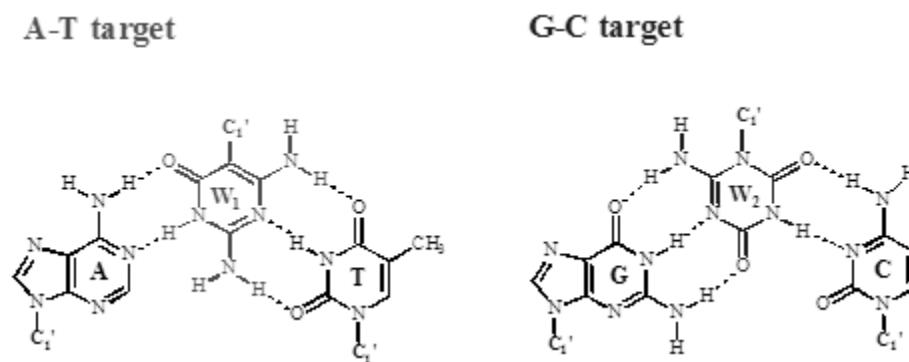
- Backbone- For this approach we need a backbone that can approach the DNA duplex and invade it. As we have discussed, because of their neutral backbone and ability of invading duplex, PNA is the first choice as backbone of these agents. Other backbones will also be considered, but our initial focus is the PNA oligomer. Other possible backbones include the methylphosphonate linkage (linking 2'-deoxynucleosides) that should limit charge-charge effects and may facilitate strand invasion.
- Unnatural nucleobase- The base should have two hydrogen bond forming faces with proper hydrogen donor-acceptor profile. This cannot be achieved by the natural bases. So we need to design these bases. Generation 1 bases, developed in our laboratory, are pyrimidine-like and Generation 2 bases are purine-like.

The PNA strand competes with and displaces the pyrimidine strand at oligopurine sites to form a D-loop structure consisting of a PNA-DNA (purine) duplex with a displaced DNA (pyrimidine) strand (Figure 4.4)<sup>14</sup>. A second PNA then binds to this initially formed duplex and results in a PNA-DNA-PNA triplex. We expect that the PNA Jw strand also binds by a strand invasion mechanism since it is designed to be complementary to the target residues, but since the second face of the Jw residue is complementary to the displaced strand, the Jw triplex results. This process can be further explored with a target sequence of oligo A-oligo T within a native duplex. The dA-dT base pairs have only 2 hydrogen bonds to disrupt, and they are more susceptible to base pair disruption<sup>15</sup>.



**Figure 4.4.** Strand invasion by a PNA to form initially the D-Loop structure, followed by a PNA-DNA-PNA triplex, or Jw triplex, depending upon the choice of the PNA strand.

Previously, our lab prepared the Generation 1 Janus-Wedge (Jw) residues. One of them is called  $W_1$  and it targets A-T and T-A<sup>16</sup>, whereas other is called  $W_2$ <sup>17</sup> and it targets G-C and C-G. The designs of these residues are presented in Figure 4.5.



**Figure 4.5.** Design of Generation 1 Jw residues. Left-  $W_1$  to target A-T and T-A; Right-  $W_2$  to target G-C and C-G.

These bases were introduced in DNA and biophysical studies were done. The results were encouraging. Below are the results found from those studies:

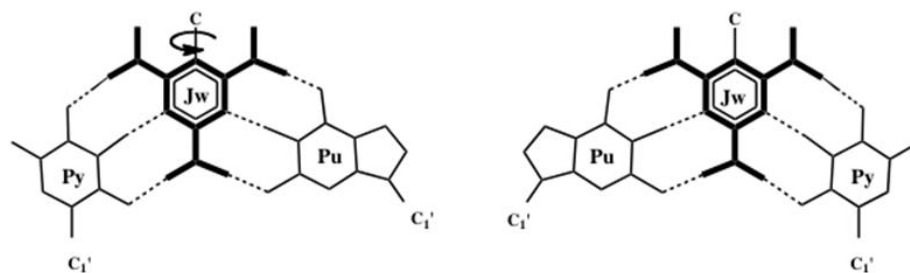
- **Thermal denaturation studies:** The thermal denaturation studies showed that these Jw triplexes showed two transitions. The first transition corresponds to the

transition from triplex structure to DNA plus PNA. The second transition accounted for the melting of DNA duplex to two single-stranded random coil transitions.

- Gel electrophoresis studies- These studies showed that the PNA can form the triplex and the mobility of the triplexes is slower than the DNA duplexes. Titration studies can be done to determine the  $K_d$  values of these bindings.
- Orientation of PNA in binding- It can be determined by tethering a phenanthroline residue to the N-terminus followed by the addition of  $\text{CuSO}_4$  and a thiol that results in cleavage of DNA<sup>18</sup>. The general conclusion was the PNA binds anti-parallel with respect to the DNA strand to which it binds stronger than the other strand.

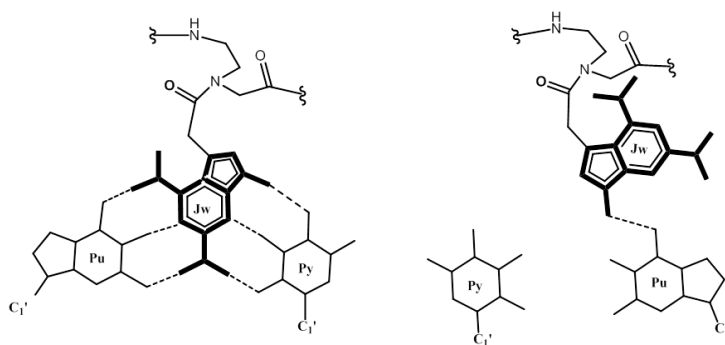
#### **4.3. Limitation with pyrimidine-like Jw triplexes**

If the two Jw residues  $W_1$  and  $W_2$  are each capable of recognizing two base pair targets, then in principle we can form triplexes at virtually any target sequence. The disadvantage to using such a format is the reduced ability to selectively target a single binding site within a large background of DNA sequences (since this motif may simply be unable to differentiate, for example dA-dT and dT-dA). This is illustrated in Figure 4.6. These Jw residues can, in principle, recognize only two base pairs since an  $180^\circ$  rotation permits one residue to recognize G-C/C-G and the second to recognize A-T/T-A base pairs, although in practice these derivatives may bind with higher affinity to one of the A-T/T-A targets (and G-C/C-G targets).



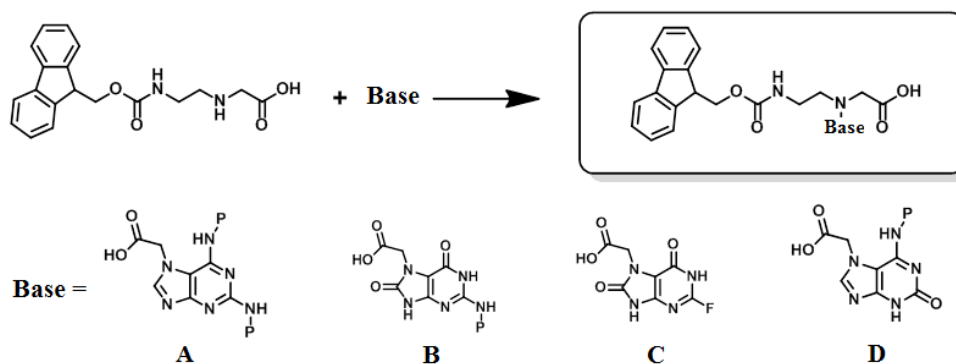
**Figure 4.6.** Specificity limitation using pyrimidine-like Jw triplexes.

In order to eliminate this targeting disadvantage, we need four Jw residues for the targeting of each of the four base pairs. Selective base pair targeting will rely upon the use of a purine-like of base for the Jw residue (Figure 4.7). Since purines have a strong preference for the anti-conformation, a given purine residue as a nucleoside should exhibit preferential binding at only one of two possible base pair targets; it would need to undergo an anti- to syn- conformational change – an unlikely event – in order to target the second base pair. This difference is likely not as strong in the PNA oligomers, although an anti-conformation, in which the bulk of the purine heterocycle is oriented away from the backbone, may still be preferred. In any case, the use of the N7-alkylated PNA monomers, and the inherent geometry of the N7 vs. N9 alkylated products will ensure that competing Watson-Crick interactions are largely absent.



**Figure 4.7.** Specificity may be achieved by using purine-like Jw residues.

With this rationale, we designed four purine-like residues retaining all the required conditions; i.e. with PNA as backbone and bases with correct hydrogen bond forming profile (Scheme 4.1). Residue A, B, C and D were envisioned to bind T-A, A-T, C-G and G-C respectively. We decided to prepare the modified bases separately with appropriate protecting groups followed by attaching to the PNA backbone (Scheme 4.1).

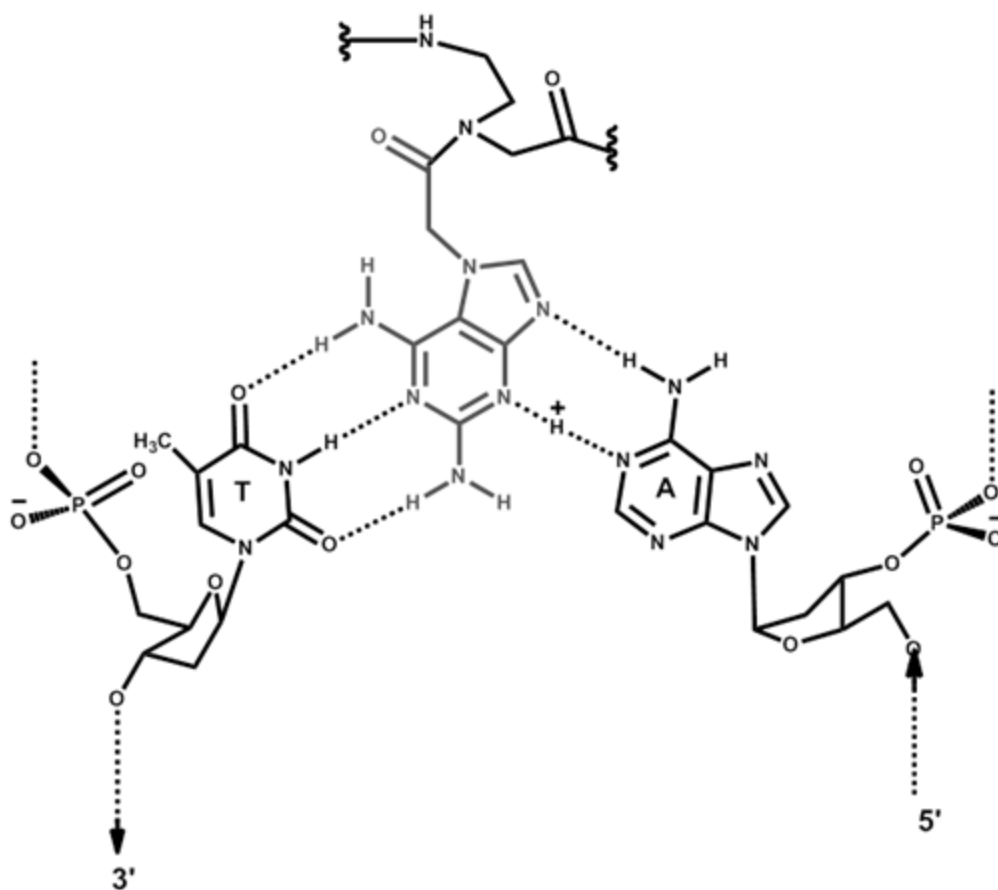


Scheme 4.1. The common strategy to synthesize the PNA monomers.

#### 4.4. Preparation of Monomer to target T-A base pair (Y1)<sup>\*</sup>

The Jw triplex for targeting T-A is presented in Figure 4.8. This derivative could form four hydrogen bonds as a triplet. An additional hydrogen bond could be possible by protonation – the proton shared between the N3-nitrogen of the Jw residue and the N1-nitrogen of the target dA residue. The pKa of the N1-nitrogen of dA has been estimated as 3.9, but this value should increase with the proximity of second protonatable nitrogen in the triplex complex – the N3-nitrogen of the Jw residue. The presence of two exocyclic amino groups and no carbonyls may also assist in raising the apparent pKa of the N3-nitrogen of this Jw residue.

<sup>\*</sup> The compound numbers are reset here for the rest of the Chapter 4.

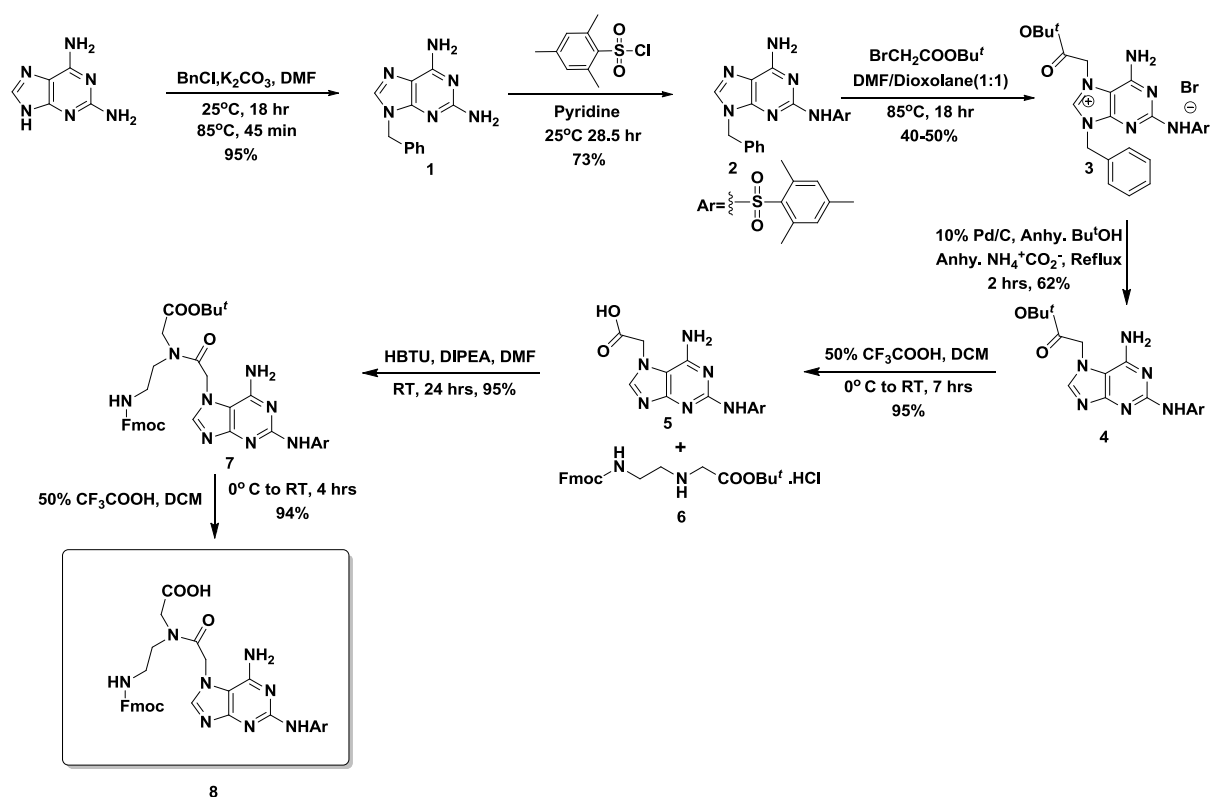


**Figure 4.8.** The Jw triplet for the T-A target.

The preparation of the monomer (Scheme 4.2) started with 2,6-diaminopurine by following the literature<sup>19</sup> under basic condition to yield 9-benzyl-2,6-diaminopurine (**1**). The reaction was clean and we got only the desired regioisomer. This was confirmed by the crystal structure (Figure 4.9A). Attempts were made to protect the two amino groups at this stage with Cbz-group. For this, CbzCl was used in different combinations of solvents, temperature and time. We had little to no success in protecting the amino groups. A more reactive alternative, Cbz-imidazole was tested. The yields were inconsistent. The maximum yield of mono-protected product that we obtained was only 17%. We never obtained the di-protected product. This led us to investigate whether



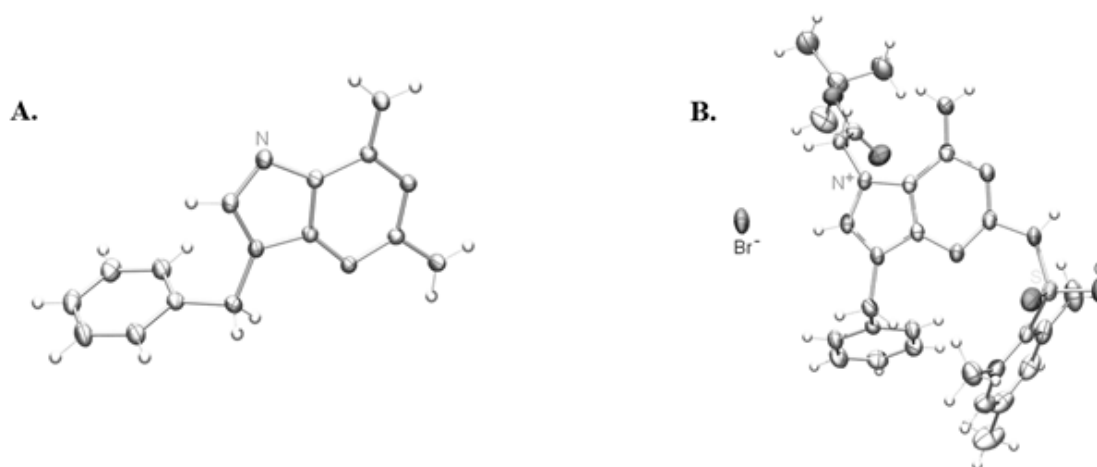
these amino-groups are unreactive enough under peptide synthesis conditions to interfere during the PNA synthesis. So we reacted compound **1** with the PNA backbone in the presence of HBTU and DIPEA in DMF. The reaction clearly showed that it was reacting. It proved that we have to protect the amino-groups efficient to move forward with the synthesis. The next choice was using mesitylenesulfonyl (Mts) group because this group has been used before in our lab for the synthesis of PNA containing Generation 1, W2 residue. The yield of this reaction was encouraging. Further optimization of this step led us to realize that adding the solution of MtsCl in DCM slowly to the solution of compound **1** in pyridine gave the best and consistent yield of mono-protected product (**2**). It was difficult to determine whether C2 or C6-amino was getting protected. It was later established that the structure was indeed compound **2**.



**Scheme 4.2:** Synthetic scheme of Y1.

It was a little puzzling that we could not detect any di-protected product. So once again we performed peptide coupling reaction of compound **2** with the PNA backbone. No reaction was observed at all which led us to believe that the C6-amino group was unreactive and will not interfere using the PNA synthesis. As a result, further attempts to obtain di-protected product were not pursued.

According to our design, we needed to install the linker at N7 of compound **2**. To do that compound **2** was reacted with  $\text{BrCH}_2\text{COOBu}^t$ . It was a difficult step because we were trying to make a highly polar salt. When the reaction was carried out in DMF at elevated temperature (not more than 85 °C), we obtained 30-32% of the product (**3**) consistently. Heating any higher resulted in decomposition of the material. Mixtures of solvents were investigated. The best possible yield (52%) was obtained in 1:1 mixture of DMF and dioxolane. The yields still vary. We think the reason is the difficulty in the isolation of the product. We were also able to obtain the crystal structure of the molecule (Figure 4.9B). This also settled the regioselectivity of the mono-protected compound (**2**).



**Figure 4.9.** Crystal structure of compound **1** (A) and **3** (B).

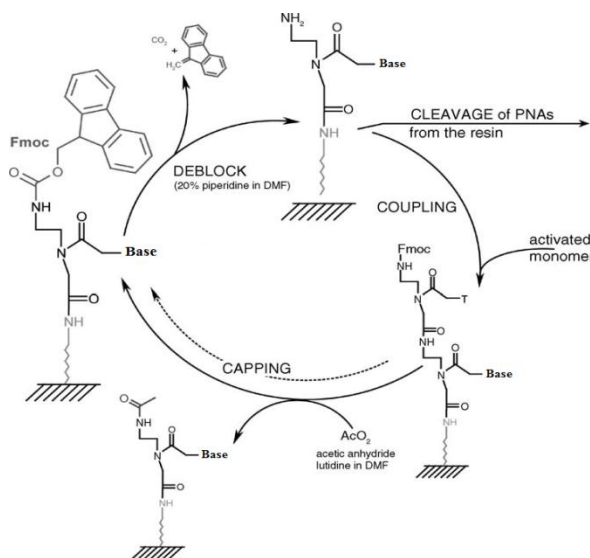
The next step was to remove the benzoyl group at the N9-position. One of the ways to remove a benzyl group from an amine is a hydrogenation reaction in the presence of Pd/C. In the first attempt, the hydrogen gas was bubbled at 1 atmospheric pressure through a solution of compound **3** in mixture of 1:1 methanol and water at 50 °C<sup>20</sup>. A trace amount of desired product was obtained, while most of the starting material was recovered. Higher pressure of hydrogen gas and higher temperature were used. We observed no considerable improvement. Hydrogenation can also be achieved by *in situ* hydrogen gas generation<sup>21</sup>. Following this methodology we were able to obtain 48-52% yield. Anhydrous methanol was used as the solvent. It is to be noted that if the reaction conditions were not anhydrous then the yield decreases significantly. The reaction also produced 10-15% methyl ester (instead of tert-butyl ester of compound **4**). Changing the solvent to tert-butanol actually improved the yield to around 70%.

To attach the modified residue to the PNA backbone, the butyl-ester (**4**) was hydrolyzed to its corresponding acid (**5**) using TFA. Coupling the acid (**5**) with the backbone was carried out employing HBTU (or HATU) and DIPEA in DMF in excellent yield. This is followed by the removal of the tert-butyl ester of the backbone to yield compound **8**. This molecule was ready to be used as monomer to synthesize the PNA.

#### **4.5. PNA Synthesis incorporating Y1**

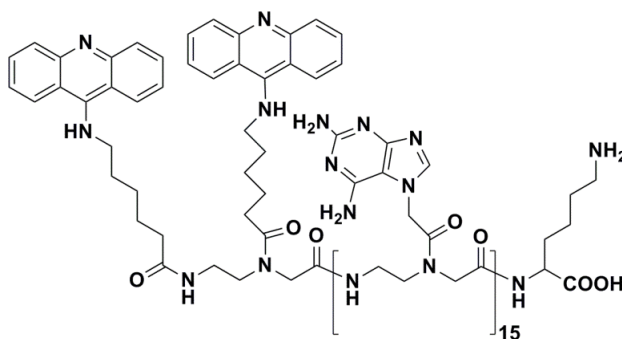
Synthesis of PNA oligomers has been carried out using a variety of monomers and coupling conditions<sup>22</sup>. Yields of oligomers depend on several factors, including the type of activator, the reaction time, the monomer excess, the pre-activation time, the times the coupling is repeated, the length of the oligomers and the temperature<sup>23</sup>. Syntheses of

sequences with a high content of purines often give poor yields due to the difficulty of coupling hindered monomers. The synthesis by Fmoc chemistry of PNA oligomers relies on the protocol developed by Egholm and Casale for the automated synthesis of PNA oligomers. In this protocol 5 equivalent of PNA monomer is initially pre-activated with HATU (4 equivalent) in the presence of the bases DIPEA (5 equivalent) and 2,6 lutidine (7.5 equivalent) and coupled for 20 min<sup>24</sup>. Although, this protocol works well under automated solid-phase PNA synthesis, when the same procedure is used for manual synthesis it requires much more equivalent of PNA monomer (3-8 equivalent) and time (1-6 hr per coupling). One of the efficient modified protocols was found to be- 2.5 equivalent of PNA, 2.5 equivalent of HOBT/HBTU, 5 equivalent NMM, pyridine (40% of the total coupling volume) with the coupling time of 30 min<sup>25</sup>. We used this protocol to synthesize our PNA strands. Overall the PNA synthesis cycle is presented in Figure 4.10. The details of the protocol will be discussed in the experimental section of the chapter.



**Figure 4.10.** PNA synthesis cycle<sup>26</sup>.

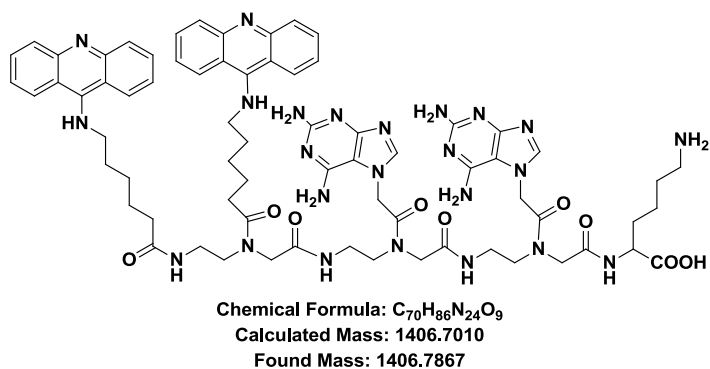
We had to perform the solid-phase synthesis manually because the solubility of our PNA monomer (**Y1**) was poor and could not be used in peptide synthesizer. The first full length PNA that we decided to prepare is presented in Figure 4.11.



**Figure 4.11.** PNA sequence with 15 **Y1** residues.

The sequence had a lysine residue at the C-terminus and two acridine groups at the N-terminus. The purpose of the lysine is to increase the solubility of PNA in water and the binding rate<sup>27</sup>. Most probably the increased cationic character of the PNA ensures a high local concentration of the PNA in proximity to the DNA. Acridines were included because they are widely used in PNAs as intercalators. They help PNA by reversibly binding to the DNA duplex. The core of the sequence consists of 15 **Y1** units to make sure the PNA binds with high stability without compromising the sequence specificity.

Since it is a challenging and time-consuming task to prepare these monomers, we decided to synthesize a model PNA which had all the necessary linkages of monomers as our desired PNA sequence (Figure 4.11). The model sequence had the lysine to **Y1** connectivity, Y1 to Y1 connectivity and Y1 to acridine connectivity (Figure 4.12).



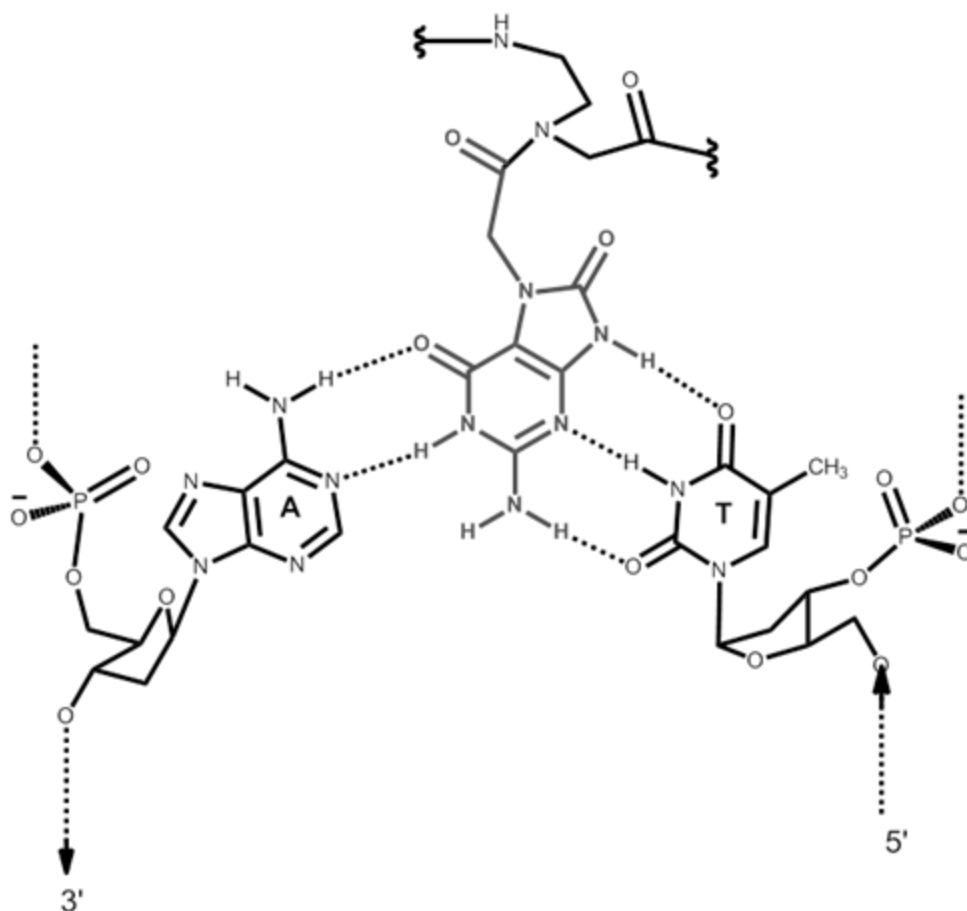
**Figure 4.12.** Model PNA sequence.

The model PNA was synthesized following the above discussed protocol. The HPLC traces were clean and mass spectrometry proved that we had our PNA. With this result in hand we decided to synthesis our desired PNA following the same protocol. We increased the coupling time from 30 min to an hour from the 8<sup>th</sup> to the 12<sup>th</sup> coupling. For the rest of the coupling, the coupling time was 2.5 hr.

The crude PNA was purified by reverse phase HPLC. Mass spectrometric analysis showed that there were peaks for various multiply-charged species of our sequence. But it also had multiply charged peaks for a species that has a molecular mass about 98 units more than our sequence. At this point we are trying to find proper conditions to separate these two species and characterize them. Once we can achieve it, we can move forward and perform binding studies with DNA duplexes.

#### 4.6. Preparation of Modified Base to target A-T base pair

The molecule structurally is N7-alkylated 8-oxoguanine derivative. It is capable of forming five hydrogen bonds in the triplex structure (Figure 4.13).

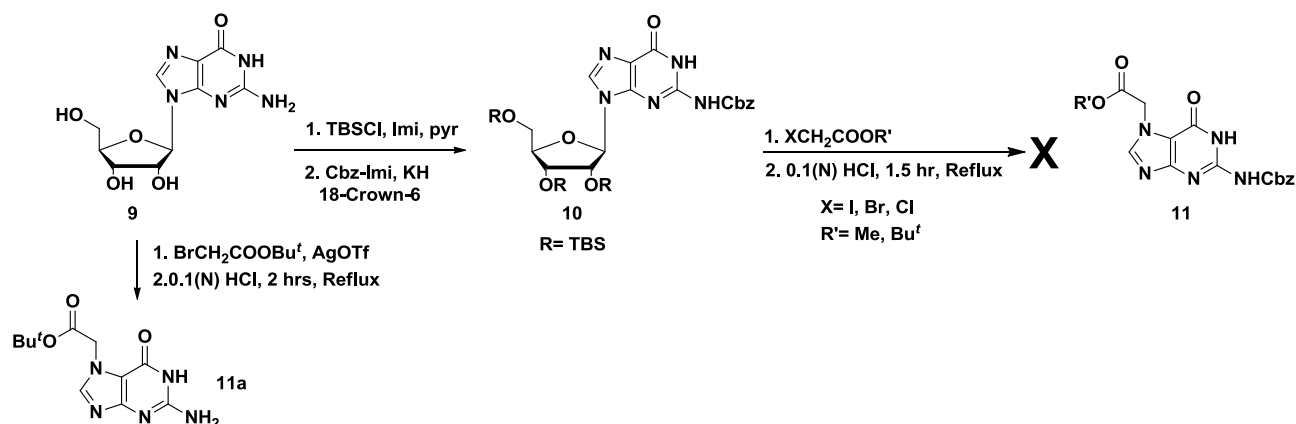


**Figure 4.13.** The Jw triplet for the T-A target.

Multiple routes were tried to synthesize this molecule. Some of them failed and some of them had very polar intermediates that were difficult to work with. We had issues with solubility, methods of isolation and characterization of products while attempting to synthesize this monomer.

The guanosine can be converted to N9-benzyl guanine by treating the former with benzyl bromide in DMSO followed by hydrolysis using 10% aq. HCl solution<sup>28</sup>. Following this procedure, we attempted to protect the three hydroxyls of the guanosine (**9**) by TBS group followed by protection of C2-amino group by Cbz group (**10**) (Scheme 4.3).

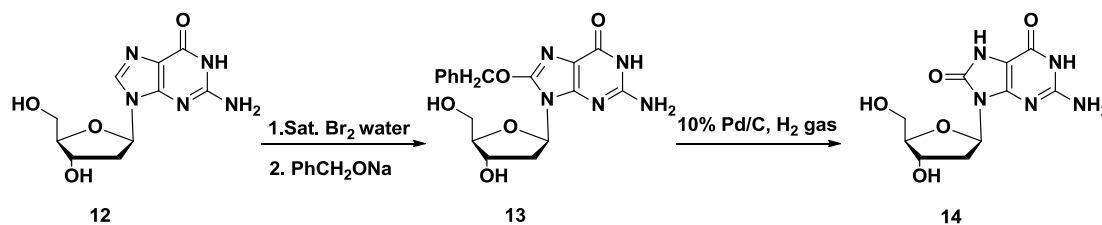
Alkylation on N7 of compound **10** was attempted by a variety of alkyl haloacetates followed by hydrolysis. The solvent, temperature and time of the reaction were screened. Unfortunately, we could not isolate the desired product. In most of the cases, the compound decomposed and could not be characterized. The only exception was when we reacted **9** with ethyl bromoacetate in the presence of AgOTf as additive. Ag<sup>+</sup> has a high affinity towards bromide and chloride. As a result Ag<sup>+</sup> increases the electrophilicity of the methylene carbon attached to bromide of ethyl bromoacetate. However, the yields were inconsistent and always below 25%. We decided to move forward using the material (**11a**) for the next step.



**Scheme 4.3.** Synthesis of the Jw base to target A-T base pair

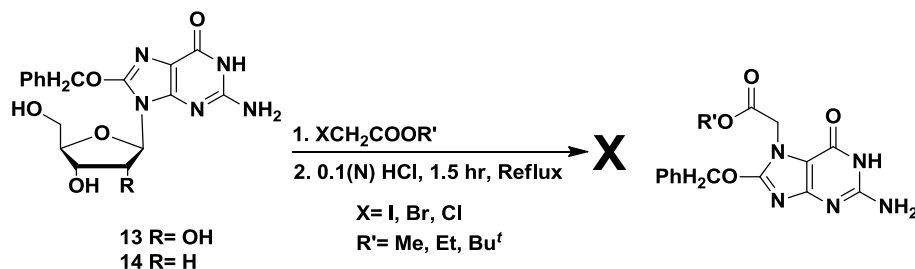
Guanosine can be converted to their 8-bromo derivative by stirring the starting material with saturated bromine water<sup>29</sup>. The bromo group can be changed to benzyloxy group by using sodium salt of benzyl alcohol, which can be further transformed to 8-oxo derivation under hydrogenation conditions<sup>30</sup> (Scheme 4.4). So we were hoping to that similar methodology can be applied to **11a**. Unfortunately, we could never brominate the 8-position of **11a**. As a result, the subsequent steps could not be explored.





**Scheme 4.4.** Methodology to prepare 7,8-dihydro-8-oxo-2'-deoxyguanosine.

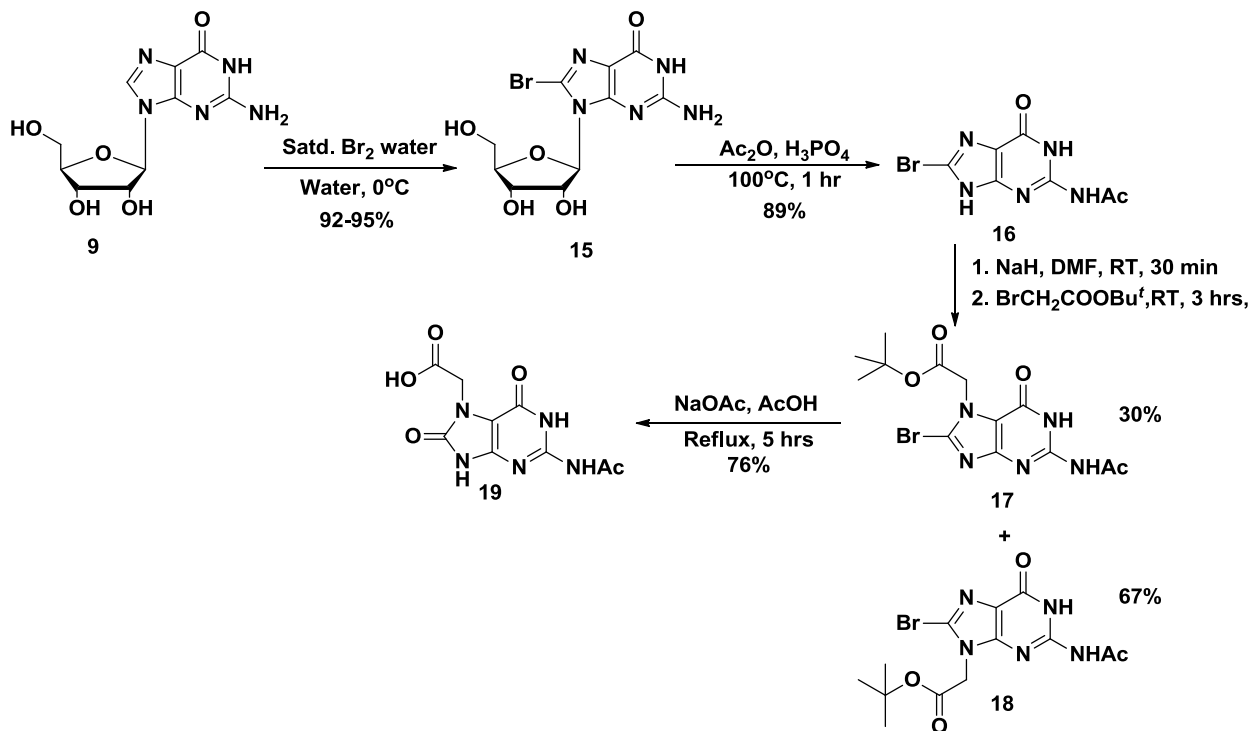
Another route was to alkylate the N7 of the 8-bromo or 8-benzyloxy derivative of guanosine or 2-deoxyguanosine. These routes were investigated in details but nothing encouraging was observed. Unfortunately, we never detected the desired product. In some cases, the glycosidic bond was cleaved, and in other cases the compound decomposed. An example of one of the routes is presented in Scheme 4.5.



**Scheme 4.5.** Attempt to alkylate N7 of 8-benzyloxy-guanosine and 8-benzyloxy-2'-deoxyguanosine.

Instead of starting with guanosine or 2-deoxyguanosine, guanine was also used as starting material. We tried to alkylate the N7 of the guanine employing different reaction conditions and reagents. Even attempts to protect the C2-amino group failed. It was soon realized that guanine is too polar to be dissolved, even partially, in the organic solvent to perform the reactions.

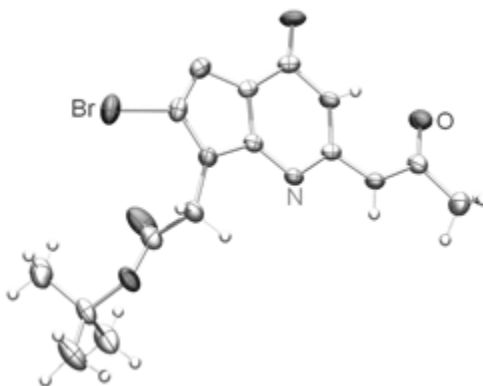
With all this information we had, we designed the Scheme 4.6 which ultimately led us to obtain the required molecule.



**Scheme 4.6.** Synthesis of the Jw base to target A-T base pair

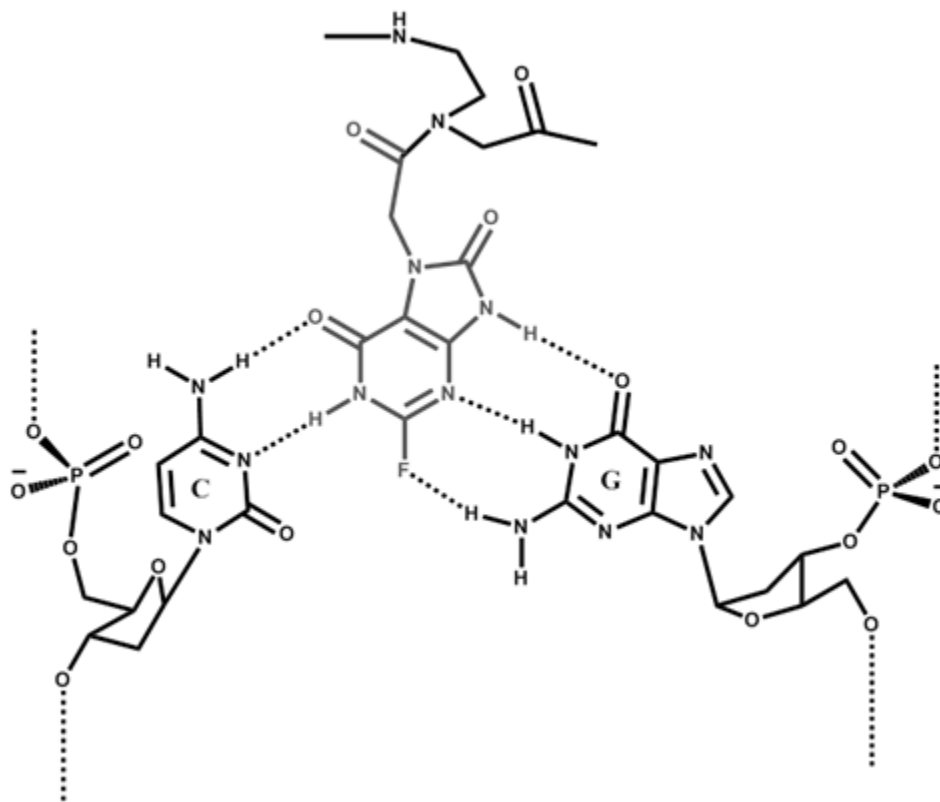
It started with the bromination of **9**. The hydroxyls can be protected by acetate followed by the cleavage of the glycosidic bond<sup>31</sup> to afford **16**. It was realized that the protecting step can be avoided by directly cleaving the glycosidic bond under the similar condition. Compound **16** was then allowed to react with tert-butyl bromoacetate. We observed that there were two very close spots on the TLC. It was expected that we will get two regioisomers at this stage. We needed the N7 regioisomer for our synthesis. But it was difficult to determine which one of the two products was the N7 since the NMR spectra were similar. Fortunately, we were able to crystallize one of the products and it turned out

to the N9 product (Figure 4.14). It was the major product as well. Nevertheless we moved forward with minor N7 product and hydrolyzed it by refluxing it with sodium acetate in acetic acid to install the 8-oxo derivative (**19**). This molecule is yet to be coupled with the PNA backbone to generate the monomer required to be used in the PNA synthesis.



**Figure 4.14.** Crystal structure of compound 18.

#### 4.7. Preparation of Modified Base to target C-G base pair

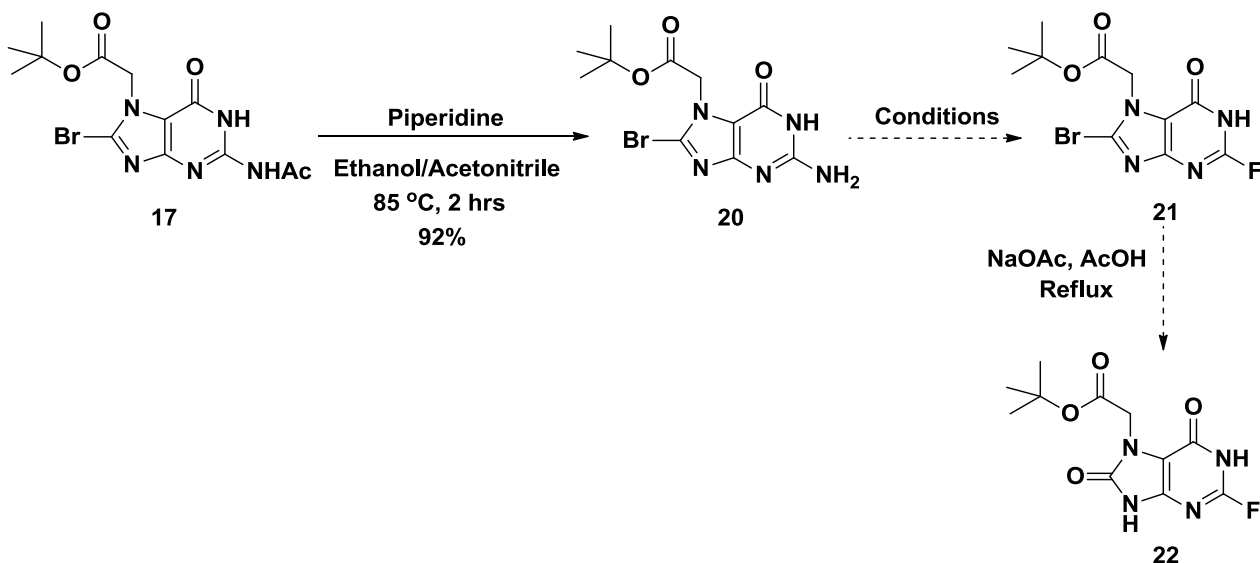


**Figure 4.15.** The Jw triplet for the C-G target.

This Jw residue should be able to form four hydrogen bonds (Figure 4.15). It has been established that organic fluorides can act as hydrogen bond acceptors<sup>32</sup>. So we expect the fifth hydrogen bond between the C2-fluorine atom of the Jw residue and the C2-amino group of the 2'-deoxyguanosine will be weak. Nevertheless, it might help targeting C-G pair over A-T base pair. From the synthetic point of view, this residue can be derived from the compound **17**.

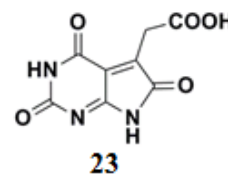
The first step is to regenerate the free C2-amino group of compound **17**. Heating compound **17** with piperidine in a 1:1 mixture of ethanol and acetonitrile yielded **20** in excellent yield. The next step is to diazotize the amino group and replace it by fluoro

group (**21**). We have not been able to do this step yet. A thorough investigation is ongoing. Once that step is standardized, the next step would be to convert 8-bromo (**21**) to 8-oxo derivative (**22**).

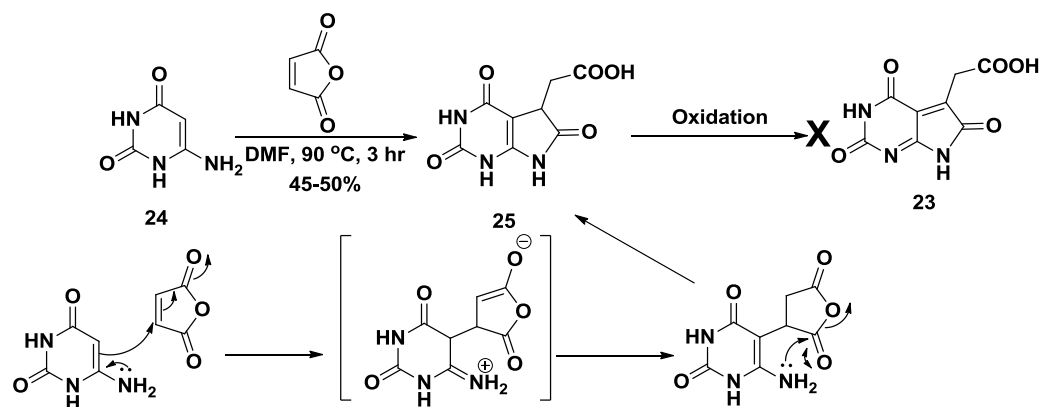


**Scheme 4.7.** Synthesis of compound **21**, Jw residue to target C-G.

An alternative Jw residue to target C-G base pair would be compound **23**. Although it is not purine-like residue but it has the correct hydrogen bond forming profile. To synthesis this molecule Scheme 4.8 was proposed.

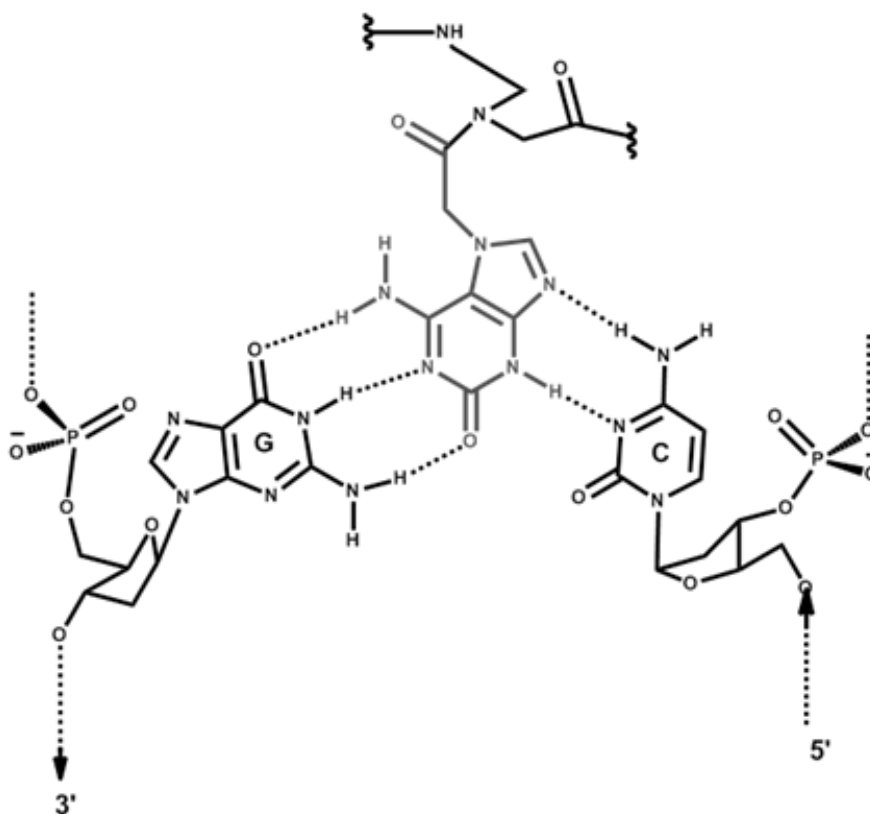


Compound **24** was heated with maleic anhydride to produce compound **24** (Scheme 4.8)<sup>33</sup>. The next step was to oxidize **25** to give us compound **23**. We hypothesized that since the oxidation of **25** to **23** will result in extension of conjugation, this reaction should be facile. A variety of oxidizing agents and conditions were tried. Unfortunately, we never observed any product. As a result, this route has not been pursued any more.



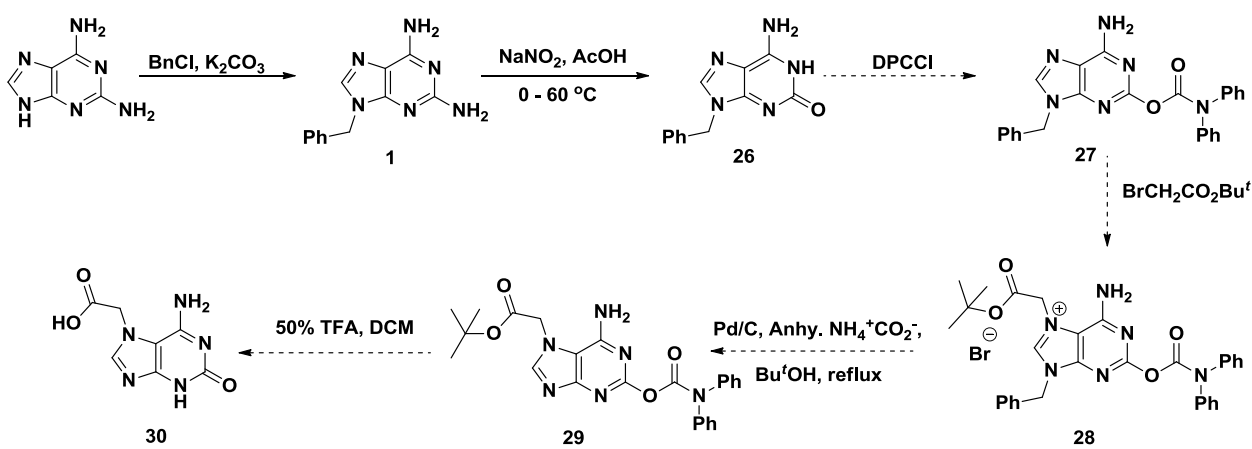
**Scheme 4.8.** Top- Attempt to synthesize compound 22. Bottom- Mechanism of first step.

#### 4.8. Attempt to Synthesize Modified Base to target G-C base pair



**Figure 4.16.** The Jw triplet for the G-C target.

Structurally the Jw residue (**30**) is N7 alkylated isoguanine derivative. The expected mode of Jw triplex formation is presented in Figure 4.16. The insolubility of isoguanine discouraged us to start the synthesis using it. We proposed Scheme 4.9 to prepare the monomer. The scheme starts with N9 benzylation of 2,6-diaminopurine (the step that we have optimized during the synthesis of **Y1**). The next step is to diazotize the C2-amino group to install the oxygen (**26**). The step has not been standardized yet. Due to pure solubility we had trouble characterizing it as well. Fortunately we were able to get the crystal structure of the molecule which confirmed the oxygen is at C2 and not C6 (Figure 4.17).



**Scheme 4.9.** Proposed route to synthesize the Jw residue (25) to target G-C

The rest of the scheme is yet to be performed. We are hopeful about it because most of the steps have been done for the synthesis of **Y1**. To decrease the polarity and the nucleophilicity of C2 oxygen of **26**, DPC can be installed (**27**). This can be followed by the alkylating the N7 of **27** to prepare the salt, **28**. This needs to be followed by the

debenzylation and hydrolysis of ester to synthesize **30**. This synthesis is under investigation and hopefully we will succeed soon.

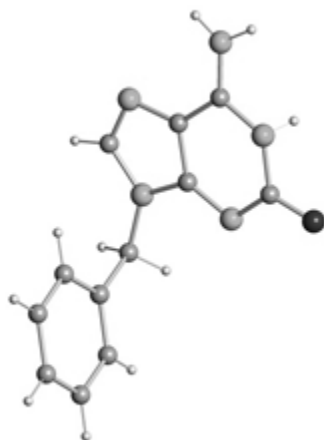


Figure 4.17. Crystal structure of compound **26**.



#### 4.9. Experimental and Methods

**Compound 2:** Compound **1** (4.0 g, 16.65 mmol) was co-evaporated with pyridine (2×50 mL). It was then dissolved in pyridine (160 mL) and mesitylene sulfonyl chloride (8.0 g, 36.63 mmol) dissolved in DCM (180 mL) was added to it using a dropping funnel. After 24 hours, there was trace amount of starting material. Another 1 equivalent of mesitylene sulfonyl chloride (3.64 g) dissolved in DCM (18 mL) was added. The reaction was completed after another 4.5 hours. The solvent was removed under high vacuum. The residue was dissolved in chloroform and washed with water three times. The organic layer was collected, dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo*. The resulting residue was purified by flash column chromatography (4.5:95.5 MeOH:DCM) to yield light yellowish white amorphous powder as compound **2** (5.14g, 73%). *R*<sub>f</sub> = 0.58 (1:9 MeOH:DCM). <sup>1</sup>H NMR (500 MHz, DMSO-D<sub>6</sub>) δ 7.97 (s, 1H), 7.28 – 7.20 (m, 3H), 7.04 – 7.00 (m, 2H), 6.88 (s, 2H), 5.10 (s, 2H), 2.64 (s, 6H), 2.18 (s, 3H). <sup>13</sup>C NMR (126 MHz, DMSO-D<sub>6</sub>) δ 153.07, 150.23, 146.00, 141.10, 139.55, 138.70, 137.04, 131.16, 128.50, 127.57, 127.22, 45.53, 22.41, 20.42. HRMS (DART-TOF) calcd for [M + H]<sup>+</sup> 423.16032, found: 423.16106.

**Compound 3:** Compound **2** (4.50 g, 10.65 mmol) was dissolved in a 1:1 mixture of anhydrous DMF (100 mL) and dioxolane (100 mL). To this BrCH<sub>2</sub>COOBu<sup>t</sup> (12.6 mL, 85.20 mmol) was added dropwise. The solution was heated at 85 °C for 18 hours when the starting material was completely consumed. The solvent was removed under high vacuum. The crude material was purified by solid pack flash column chromatography (7.5:92.5 MeOH:DCM) to yield white amorphous powder as compound **3** (3.0 g, 45%). *R*<sub>f</sub> = 0.22 (1:9 MeOH:DCM). <sup>1</sup>H NMR (500 MHz, DMSO-D<sub>6</sub>) δ 11.76 (s, 1H), 9.47 (s,

1H), 7.72 (s, 2H), 7.39 – 7.32 (m, 3H), 7.15 (dt, J = 3.6, 2.7 Hz, 2H), 6.94 (t, J = 2.8 Hz, 2H), 5.46 (s, 2H), 5.32 (d, J = 6.6 Hz, 2H), 2.64 (d, J = 3.8 Hz, 6H), 2.20 (s, 3H), 1.37 (s, 9H). <sup>13</sup>C NMR (126 MHz, DMSO-D<sub>6</sub>) δ 165.09, 155.12, 152.52, 148.98, 142.05, 141.37, 138.88, 134.17, 133.93, 131.47, 128.85, 128.59, 127.90, 106.45, 83.34, 50.24, 47.47, 27.47, 22.26, 20.47. HRMS (ESI-TOF) calcd for [M] 537.22785, found: 537.22900.

**Compound 4:** Compound **3** (2.50 g, 4.05 mmol), ammonium formate (1.28 g, 20.25 mmol) and 1% Pd/C (2.50 g) were taken in tert-butanol (40 mL). The mixture was refluxed for 2 hr. The TLC indicated that the starting material was consumed. The solution was filtered hot and washed twice by hot methanol. The filtrate was allowed to cool down to room temperature. The solvent was removed under high vacuum. The crude material was purified by flash column chromatography (4:96 MeOH:DCM) to yield white amorphous powder as compound **4** (1.12g, 62%). *R<sub>f</sub>* = 0.47 (1:9 MeOH:DCM). <sup>1</sup>H NMR (500 MHz, DMSO-D<sub>6</sub>) δ 10.81 (s, 1H), 8.02 (s, 1H), 6.94 (s, 2H), 6.51 (s, 1H), 5.21 (s, 2H), 2.70 (s, 6H), 2.21 (s, 3H), 1.36 (s, 9H). <sup>13</sup>C NMR (126 MHz, DMSO-D<sub>6</sub>) δ 174.12, 166.03, 154.01, 150.61, 148.96, 125.06, 120.36, 104.99, 83.43, 45.83, 28.01, 24.22 (Please note that two expected peaks are missing even after 20000 scans). HRMS (DART-TOF) calcd for [M + H]<sup>+</sup> 447.18145, found: 447.18221.

**Compound 5:** Compound **4** (1.0 g, 1.30 mmol) was taken in DCM (13 mL) and cooled down to 0 °C. TFA (13 mL) was added dropwise when the solution slowly became clean and pale yellow. The temperature was allowed to warm up to the room temperature. After 7 hr, the reaction was completed. The solvent was removed and the product was precipitated out by adding diethyl ether. It was centrifuged and the solid was washed with diethyl ether twice. The white solid was characterized as product **5** (830 mg, 95%). <sup>1</sup>H

NMR (500 MHz, DMSO-D<sub>6</sub>)  $\delta$  8.16 (s, 1H), 6.94 (s, 2H), 6.68 (s, 2H), 5.23 (s, 2H), 2.70 (s, 6H), 2.21 (s, 3H). <sup>13</sup>C NMR (126 MHz, DMSO-D<sub>6</sub>)  $\delta$  170.00, 153.40, 152.15, 146.57, 141.54, 139.61, 135.83, 131.63, 108.28, 48.12, 22.93, 20.86.

Compound **7**: Compound **5** (750 mg, 1.92 mmol), **6** (755 mg, 1.75 mmol) and HBTU (796 mg, 2.10 mmol) were taken in DMF. To this DIPEA (0.67 mL, 3.85 mmol) was added. It was stirred at ambient temperature for 24 hr when the TLC indicated the completely consumption of compound **6**. The solvent was removed under high vacuum. The crude material was purified by flash column chromatography (2.5:97.5  $\rightarrow$  4.5:95.5 MeOH:DCM) to yield yellowish-white foam as compound **7** (1.40 g 95%).  $R_f$  = 0.52 (1:19 MeOH:DCM). HRMS (ESI-TOF) calcd for  $[M + H]^+$  769.312643, found: 769.31320.

Compound **8**: Compound **7** (1.0 g, 1.30 mmol) was taken in DCM (13 mL) and cooled down to 0 °C. TFA (13 mL) was added dropwise when the solution slowly became clean and pale yellow. The temperature was allowed to warm up to the room temperature. After 4 hr, the reaction was completed. The solvent was removed and the product was precipitated out by adding diethyl ether. It was centrifuged and the solid was washed with diethyl ether twice. The white solid was characterized as product **8** (870 mg, 94%). HRMS (ESI-TOF) calcd for  $[M + H]^+$  713.250042, found: 713.25010.

### ***PNA Synthesis***

The PNA molecules were synthesised on Fmoc-protected MBHA resin with a loading capacity of about 0.54 mmol/g. After swelling 70 mg of the resin for 15 min DMF in the fritted column, the coupling cycle was started with the cleavage of the Fmoc-protection

groups. This was done by two successive incubations (5 min each) with 3 ml 20% piperidine in DMF solution. The Fmoc groups and piperidine were washed away six times with 3 mL DMF followed by a double coupling (30 min each) of the first monomer. Each of the two coupling solutions consisted of 2.5 equivalent of required monomer (67 mg of PNA monomer or 44 mg of Fmoc-Lys-Boc or 60 mg of Fmoc-acrydine monomer or 29 mg of terminal acrydine monomer), 2.5 equivalent of HBTU (36 mg) and HOBt (13 mg) dissolved in 1.37 mL DMF and 150  $\mu$ L of NMM solution (0.2(M) NMM in pyridine). After coupling, the resin was washed six times with 3 mL DMF. It was followed by capping by capping solution (acetic anhydride/2,6-lutidine/DMF in 5/6/89 ratio by volume) for 5 min.. Finally the resin was washed another six times with 3 mL DMF. The cycle was repeated until the synthesis of the desired PNA sequence was completed. The terminal Fmoc group was removed as described above. The resin was washed another six times with 3 mL of DMF and ten times with 3 mL of DCM. The resin was dried under flowing nitrogen gas.

After synthesis, the final PNA molecules were cleaved from the dried resin by incubating 10 mg of resin in 100  $\mu$ L thioanisol and 50  $\mu$ L 1,2-ethanedithiol at 0 °C for 10 min. 650  $\mu$ L of TFA was added to the solution and kept for another 10 min at 0 °C. To this mixture 200  $\mu$ L trifluoromethane sulfonic acid was added and the solution was kept at 0 °C for 10 min. It was allowed to warm up to room temperature and kept for another 2 hr. The solution was filtered and the resin was washed with TFA (2 $\times$ 250  $\mu$ L). Precipitation of PNAs was done by adding 5 ml of ice-cold diethyl ether. PNAs were washed another time with ether, dried and dissolved in 20 mL water. It was then purified by reverse phase

semi-prep HPLC. Quality control of all synthesized PNAs was done by MALDI-TOF mass spectrometry.

**Compound 16:** Compound **15** (3.0 g, 8.28 mmol) was suspended in acetic anhydride (21 mL) at room temperature. To this  $\text{H}_3\text{PO}_4$  (0.16 mL) was added. The solution was heated at 100 °C for 1 hr. The solution became clear after 15 minutes and then slowly white precipitate started crushing out of the solution. After the reaction, the solution was filtered and the residue was washed with acetone (2×75 mL). The off-white residue was characterized as product **16** (2.0 g, 89%).  $^1\text{H}$  NMR (500 MHz, DMSO- $\text{D}_6$ )  $\delta$  12.03 (s, 1H), 11.62 (s, 1H), 2.16 (s, 3H).  $^{13}\text{C}$  NMR (126 MHz, DMSO- $\text{D}_6$ )  $\delta$  173.88, 153.14, 148.03, 123.83, 118.98, 24.18. HRMS (DART-TOF) calcd for  $[\text{M} + \text{H}]^+$  271.97831, found: 271.97827.

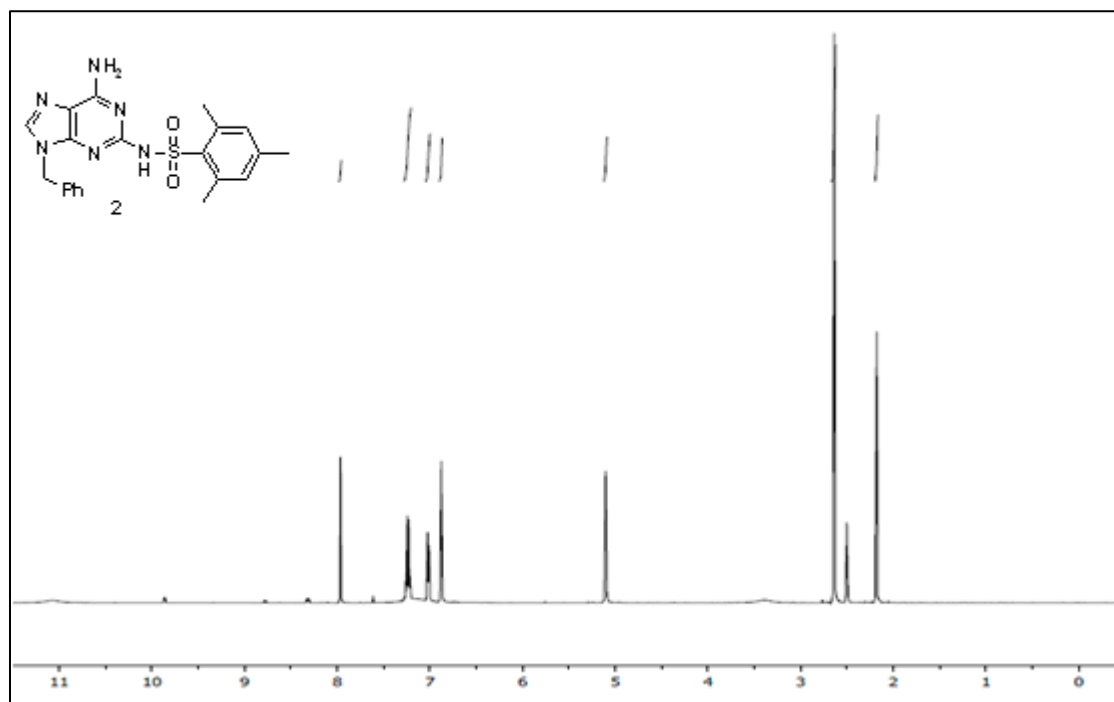
**Compound 17:** Compound **16** (1.75 g, 6.43 mmol) and sodium hydride (283 mg of 60% NaH dispersed in mineral oil, 7.07 mmol) were taken in anhydrous DMF (60 mL). The solution was stirred at room temperature for 30 min. To this  $\text{BrCH}_2\text{COOBu}^t$  (1.05 mL, 7.07 mmol) was added dropwise and the mixture was allowed to stir for 3 hr. The TLC indicated the complete consumption of starting material. The solvent was removed under high vacuum. The crude material was purified by solid pack gravity column chromatography (2.5:97.5  $\rightarrow$  3:9 MeOH:DCM) to yield white powder as required compound **17** (745 mg, 30%) ( $R_f$  = 0.22; 2.5:7.5 MeOH:DCM) and N9 product, compound **18** (1.66 g, 67%) ( $R_f$  = 0.19; 2.5:7.5 MeOH:DCM). Characterization of compound **17**:  $^1\text{H}$  NMR (500 MHz, DMSO- $\text{D}_6$ )  $\delta$  5.06 (s, 2H), 2.17 (s, 3H), 1.42 (s, 9H).  $^{13}\text{C}$  NMR (126 MHz, DMSO- $\text{D}_6$ )  $\delta$  173.48, 165.73, 156.05, 151.62, 147.58, 131.39, 113.80, 82.77, 48.19, 27.56, 23.70. HRMS (DART-TOF) calcd for  $[\text{M} + \text{H}]^+$  386.04639,

found: 386.04783. Characterization of compound **18**:  $^1\text{H}$  NMR of compound **11** (500 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  4.84 (s, 2H), 2.21 (s, 3H), 1.48 (s, 9H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  174.94, 167.02, 156.32, 152.24, 149.99, 126.75, 121.23, 84.62, 46.57, 28.16, 23.85. HRMS (DART-TOF) calcd for  $[\text{M} + \text{H}]^+$  386.04639, found: 386.04758.

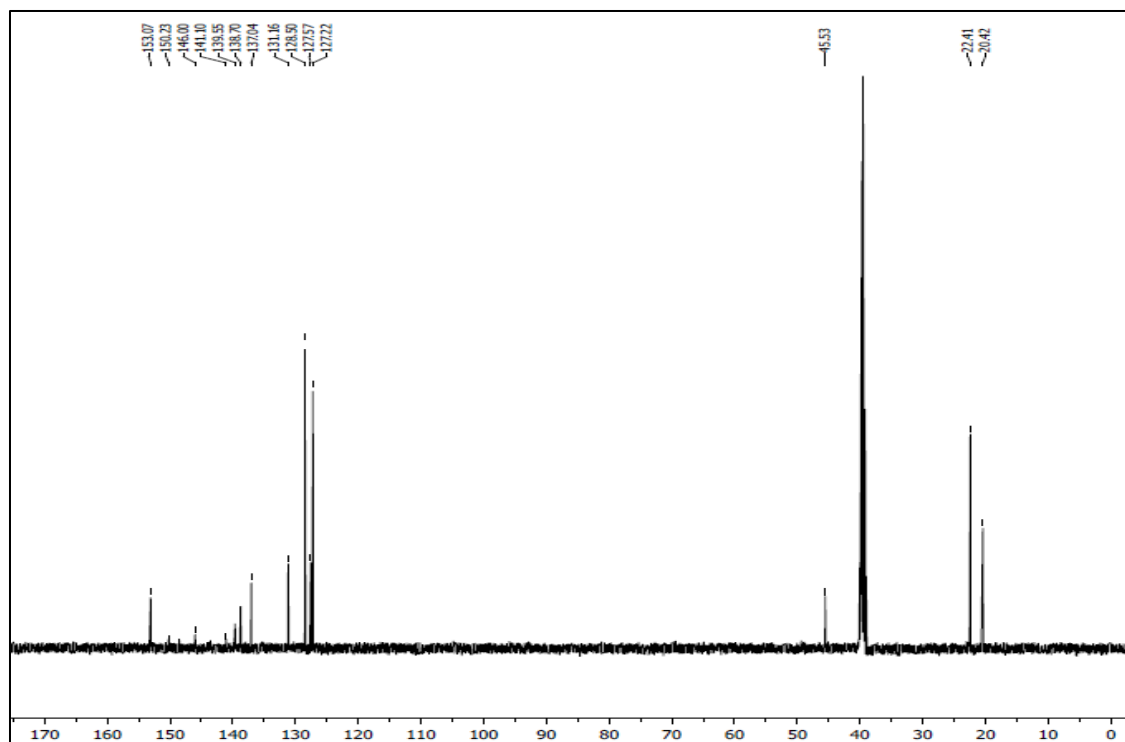
Compound **19**: Compound **17** (650 mg, 1.68 mmol) and sodium acetate (1.10 g, 13.44 mmol) were refluxed in acetic acid for 5 hr. After the reaction the solution was allowed to cool to room temperature. Acetic acid was removed under high vacuum. Denatured ethanol (150 mL) was added to the crude and it was stirred at 40 °C for 1.5 hr. Then the solution was filtered hot (2×50 mL). The white residue was characterized as compound **19** (342 mg, 76%).  $^1\text{H}$  NMR (500 MHz,  $\text{D}_2\text{O}$ )  $\delta$  4.50 (s, 2H), 2.27 (s, 3H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{D}_2\text{O}$ )  $\delta$  175.06, 174.88, 153.37, 151.08, 147.21, 145.43, 105.24, 44.93, 23.24. HRMS (DART-TOF) calcd for  $[\text{M} + \text{H}]^+$  268.06819, found: 268.06771.

Compound **20**: Compound **17** (500 mg, 1.30 mmol) was taken in 1:1 mixture of ethanol (13 mL) and acetonitrile (13 mL). To this piperidine (0.32 mL, 3.25 mmol) was added and the solution was heated at 85 °C for 2 hr. The solution was cooled to room temperature and then cooled to -20 °C overnight. The product crashed out of the solution. It was collected by filtering the solution and washing the precipitate with ethanol. The white solid was characterized as compound **20** (410 mg, 92%). HRMS (DART-TOF) calcd for  $[\text{M} + \text{H}]^+$  344.03583, found: 344.03663. I could not obtain the proton or carbon NMR spectra because of its extreme insolubility in any of the solvents or mixture of solvents.

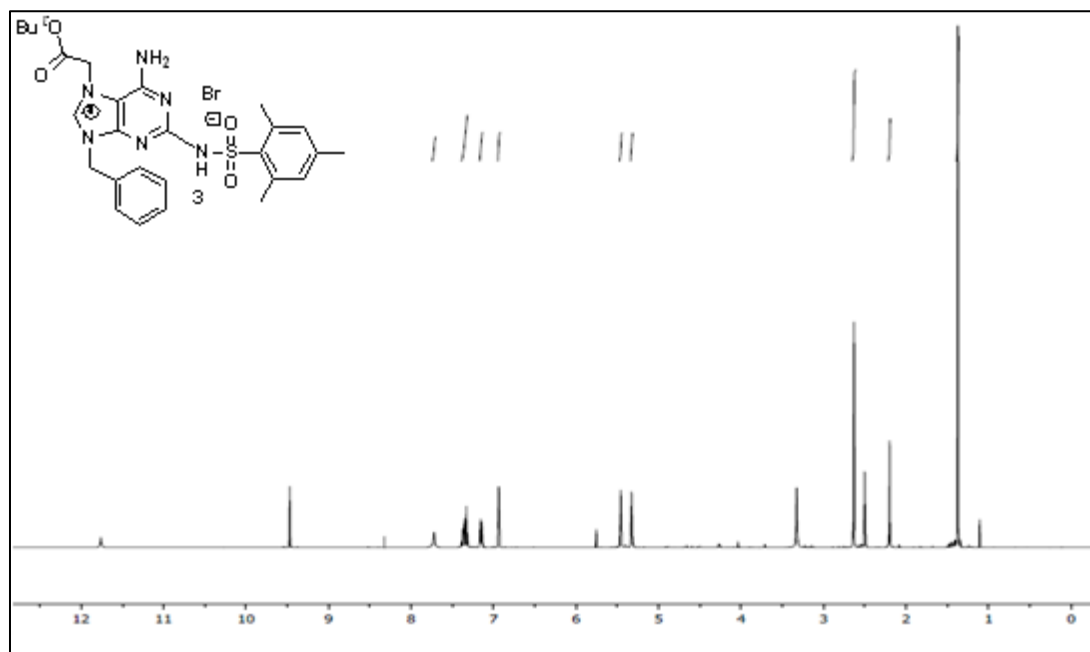
<sup>1</sup>H NMR of Compound **2**



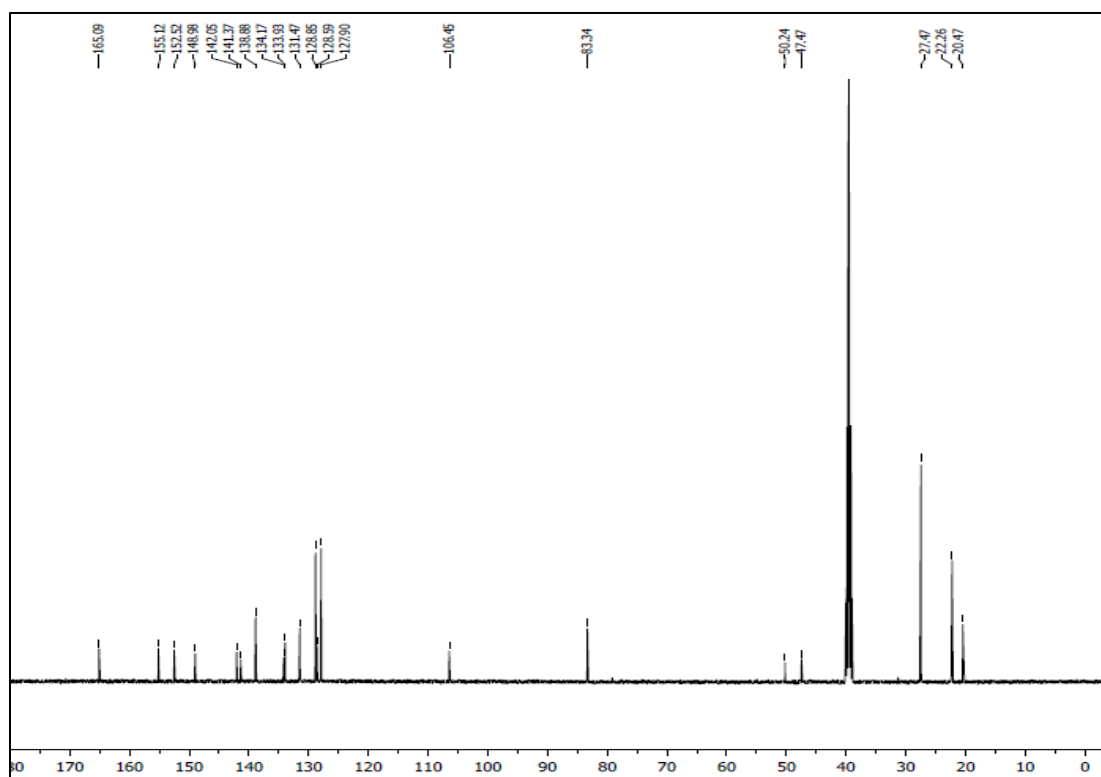
<sup>13</sup>C NMR of Compound **2**



<sup>1</sup>H NMR of Compound **3**

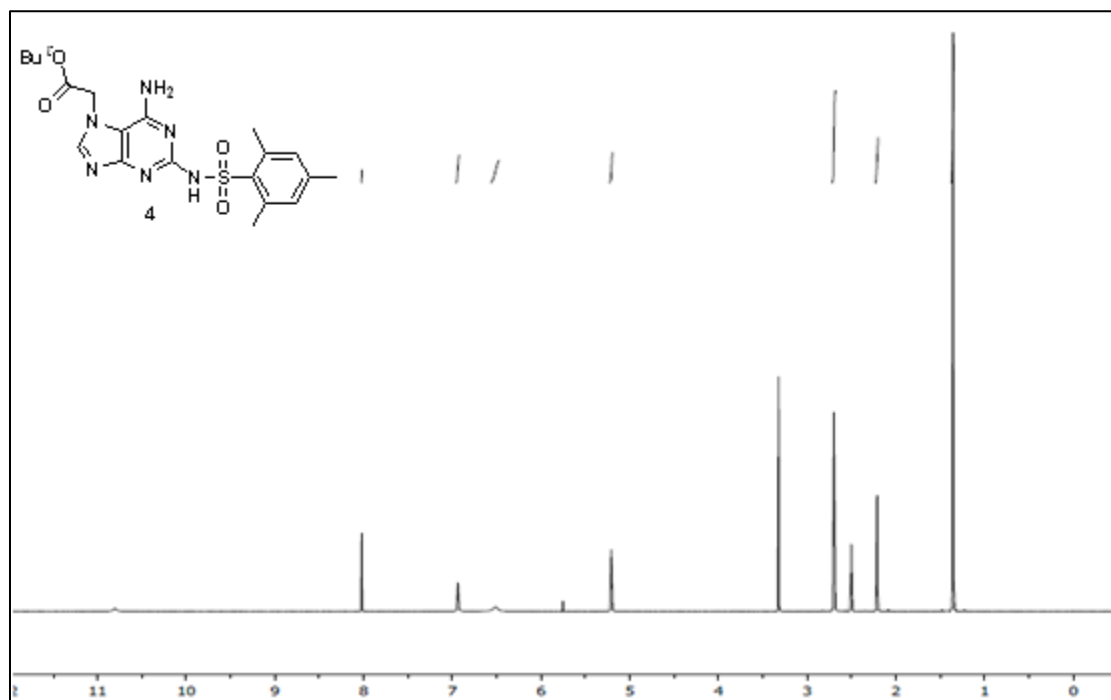


<sup>13</sup>C NMR of Compound **3**

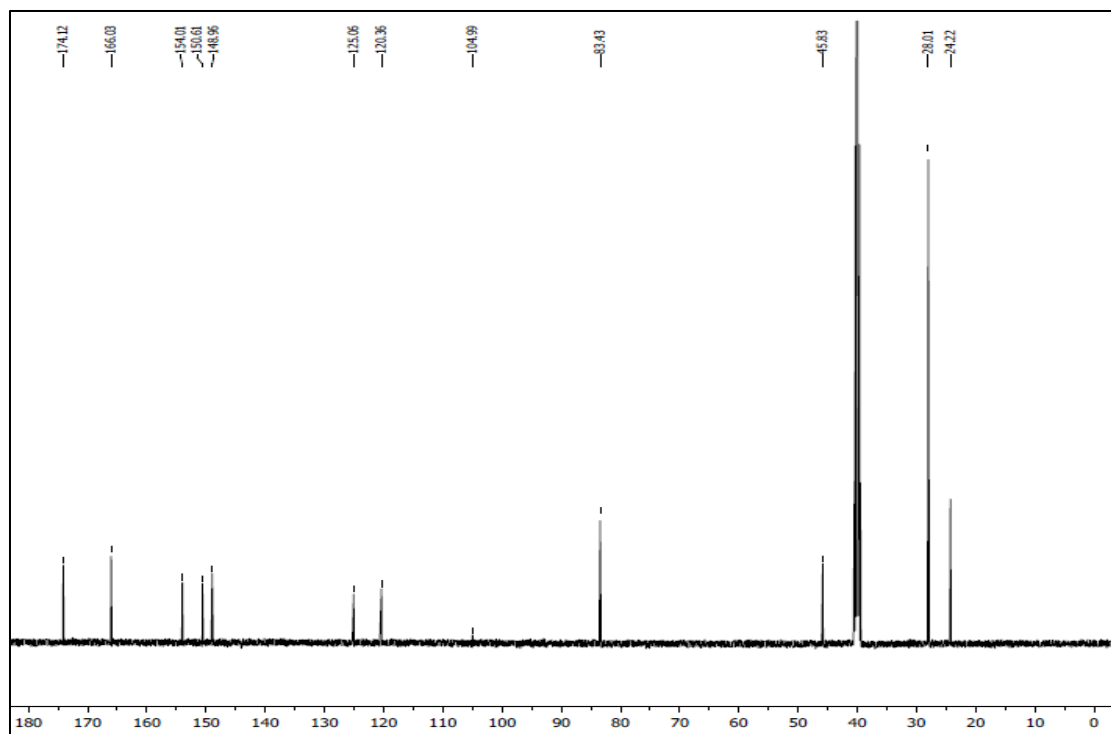




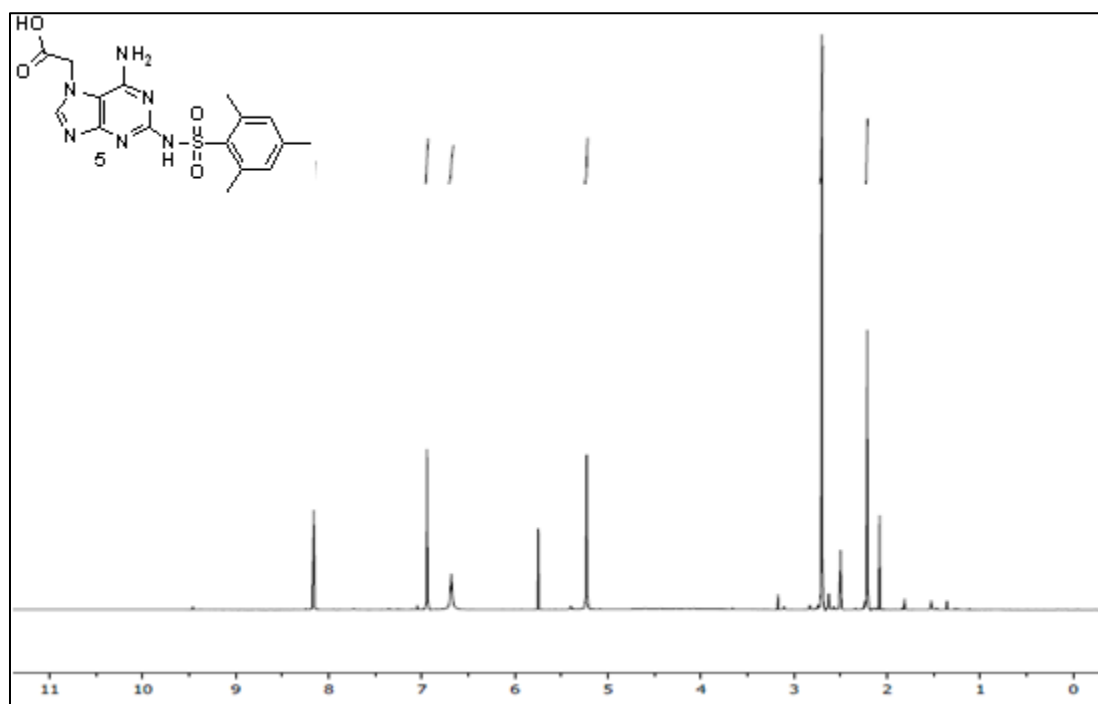
<sup>1</sup>H NMR of Compound 4



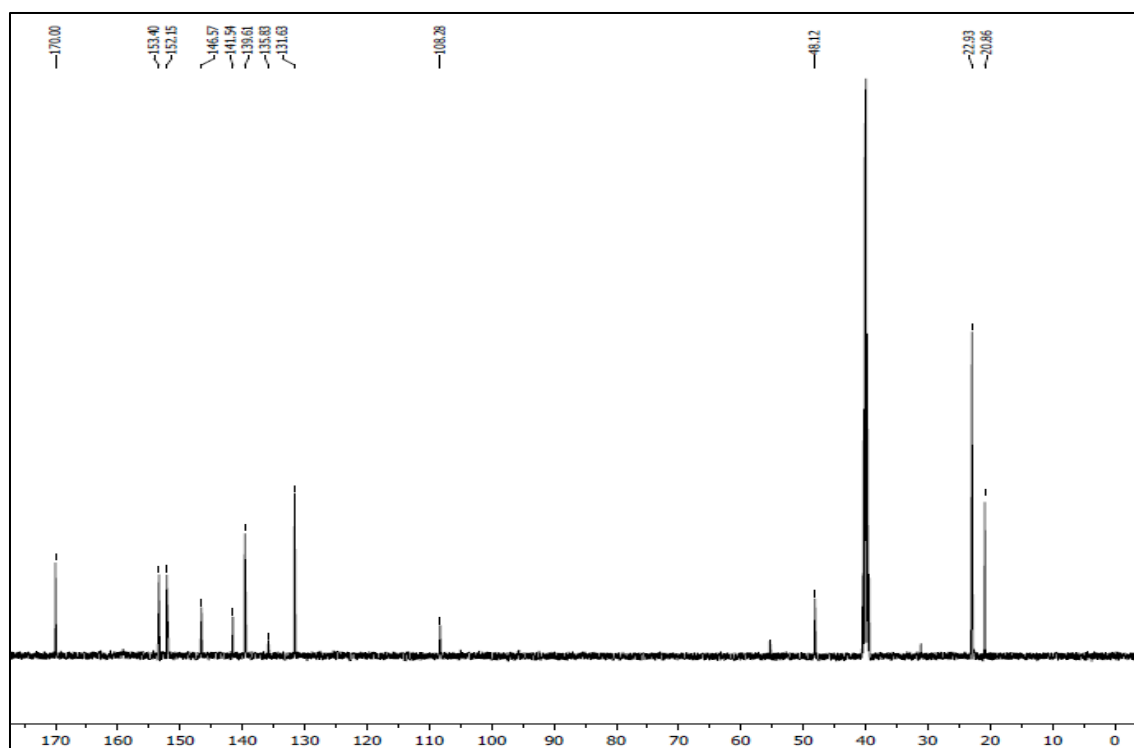
<sup>13</sup>C NMR of Compound 4



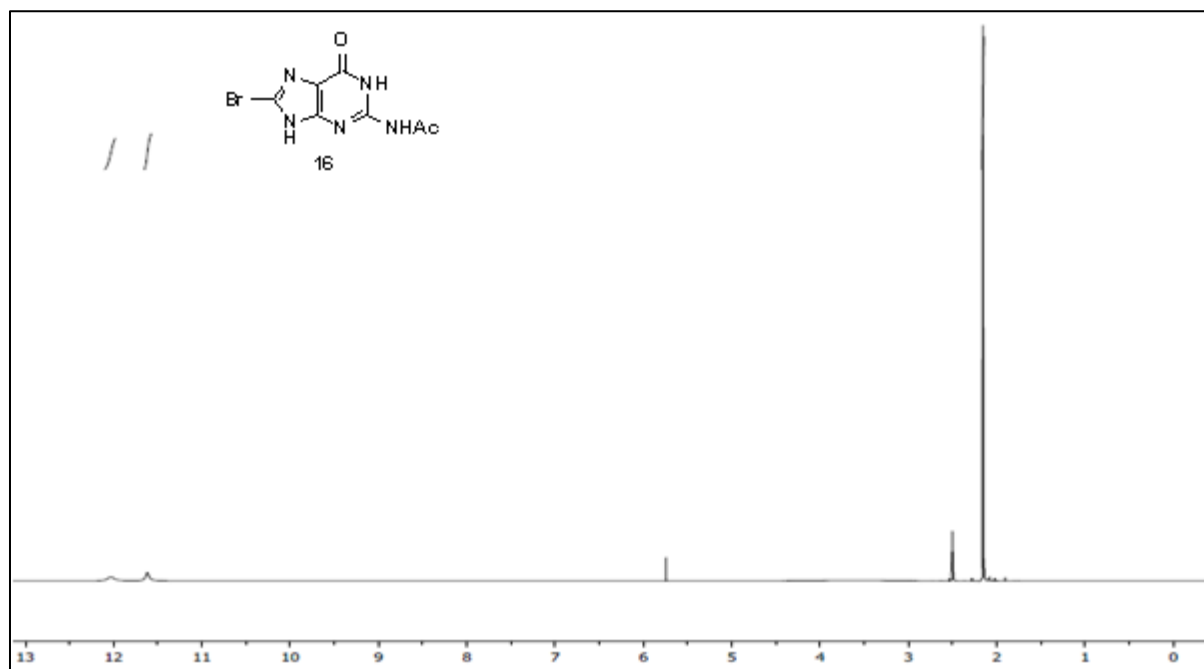
<sup>1</sup>H NMR of Compound **5**



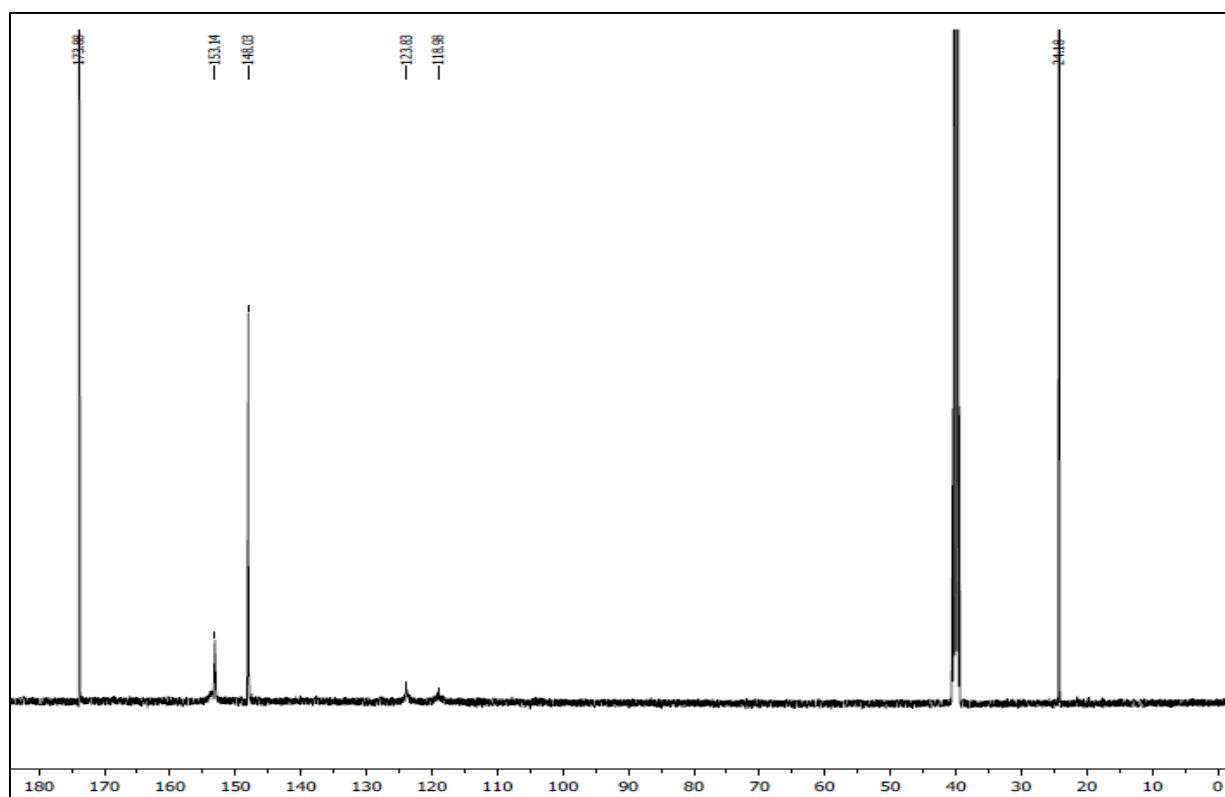
<sup>13</sup>C NMR of Compound **5**



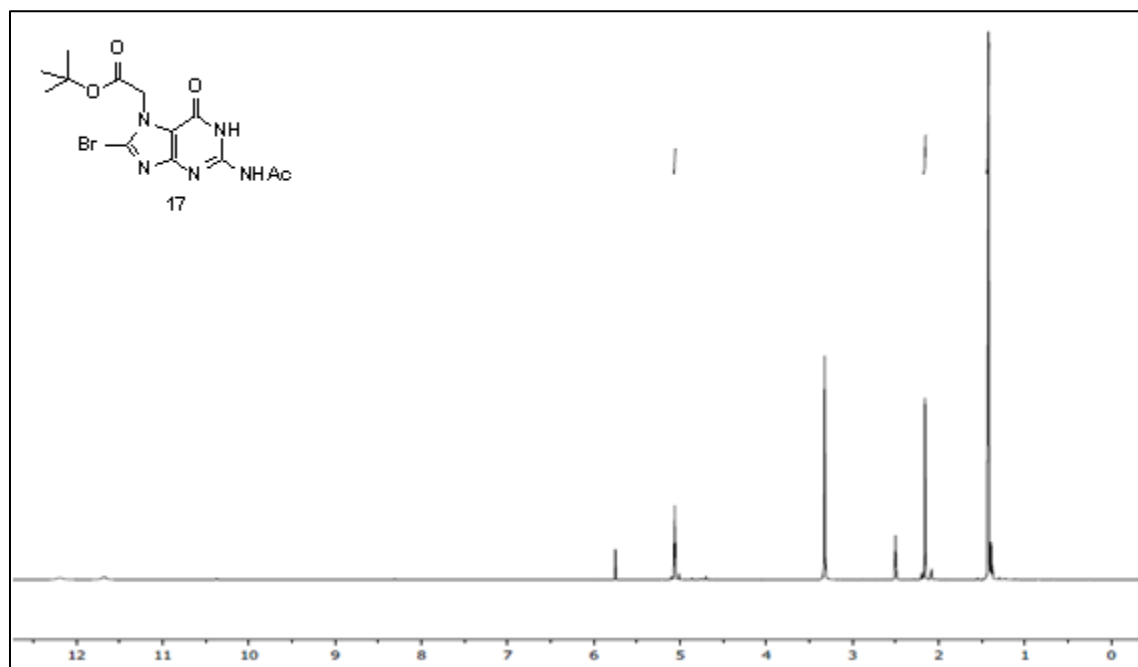
$^1\text{H}$  NMR of Compound **16**



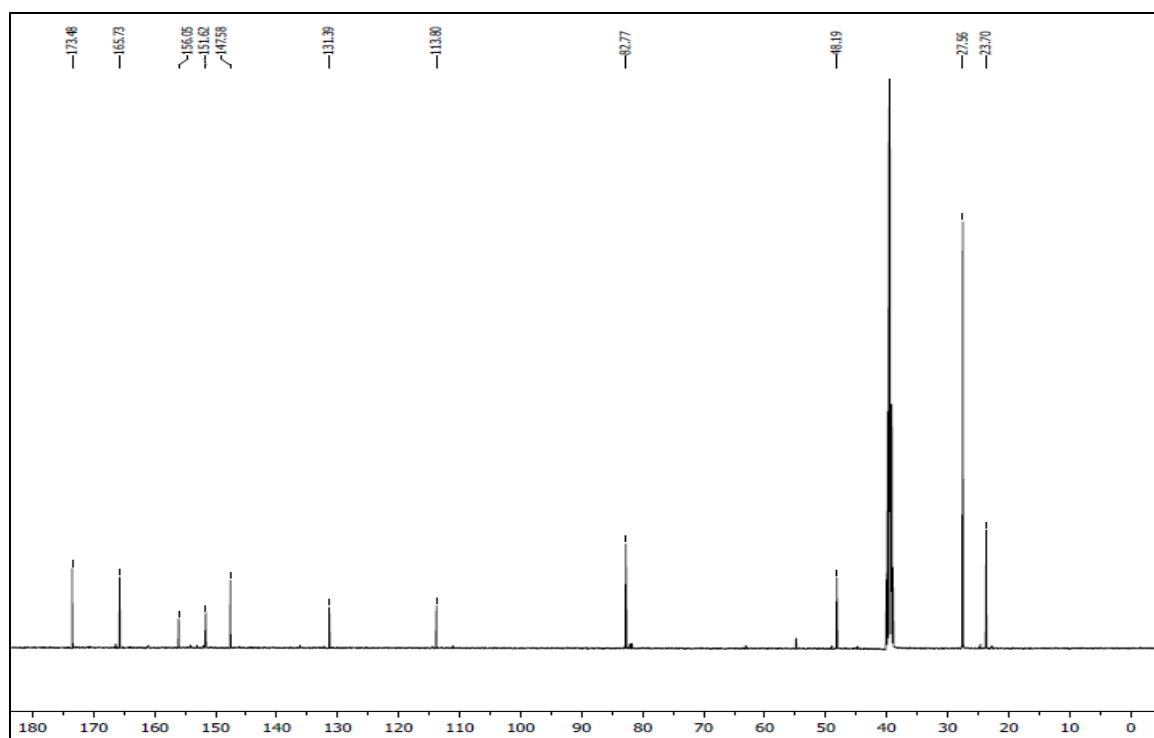
$^{13}\text{C}$  NMR spectrum of Compound **16**



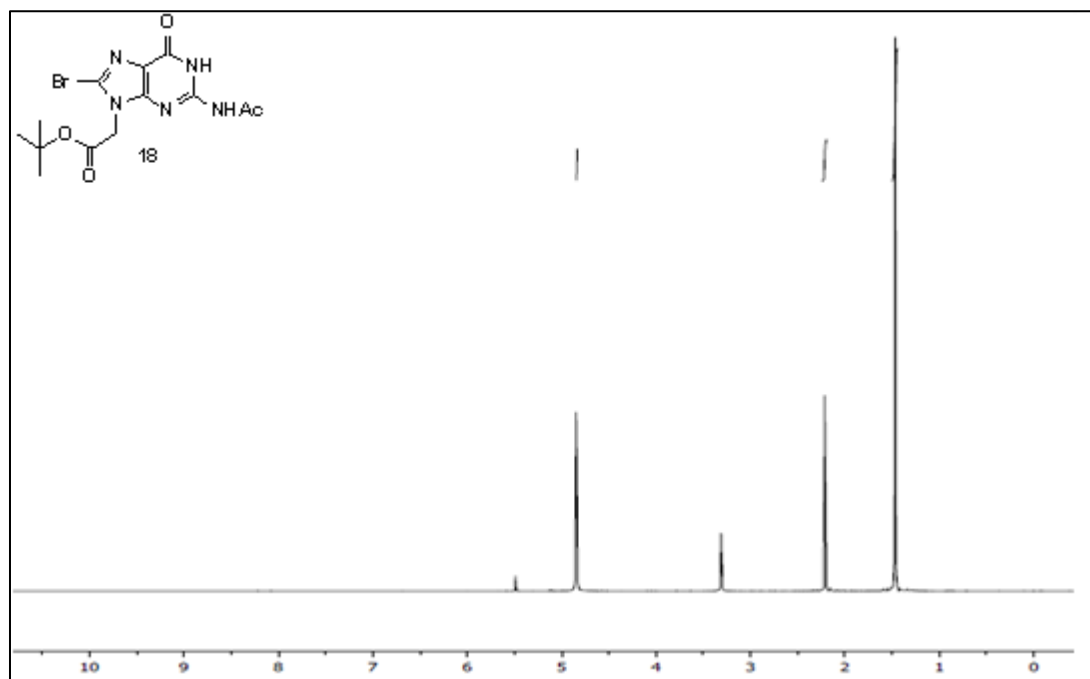
$^1\text{H}$  NMR of Compound **17**



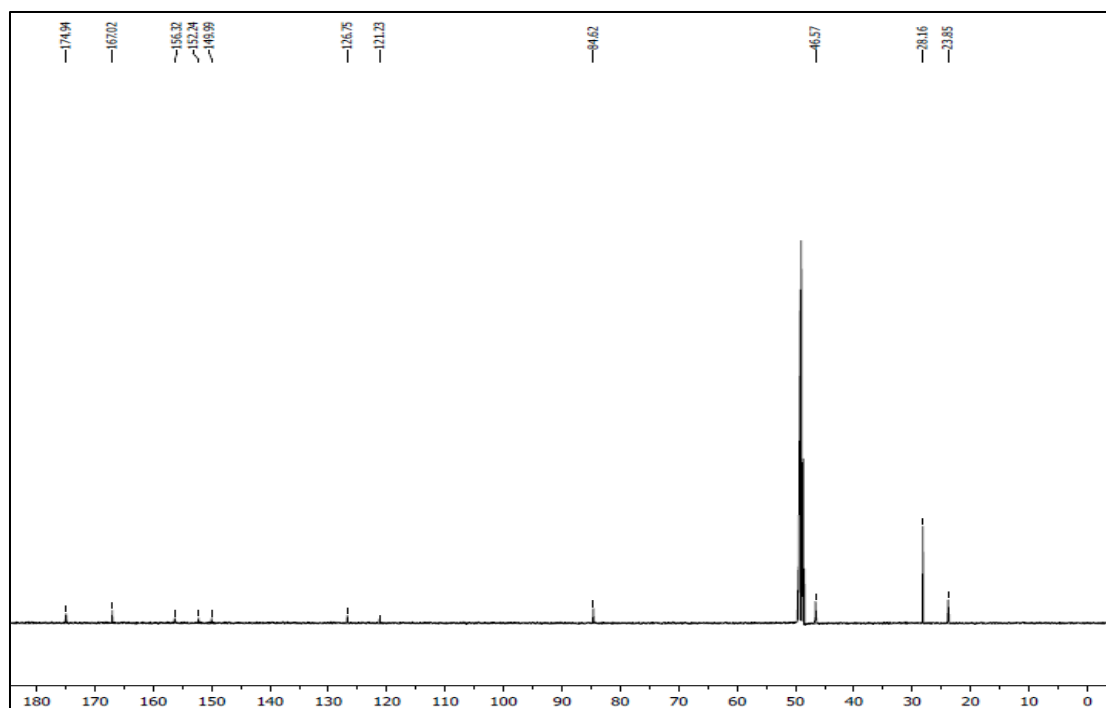
$^{13}\text{C}$  NMR of Compound **17**



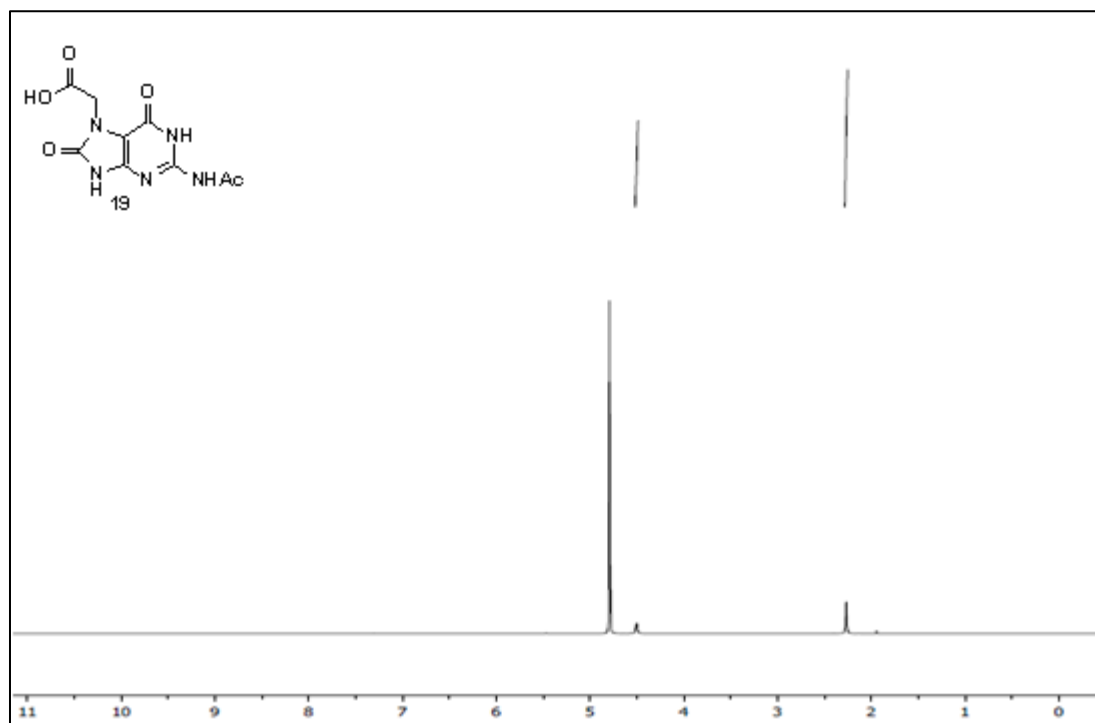
<sup>1</sup>H NMR of Compound **18**



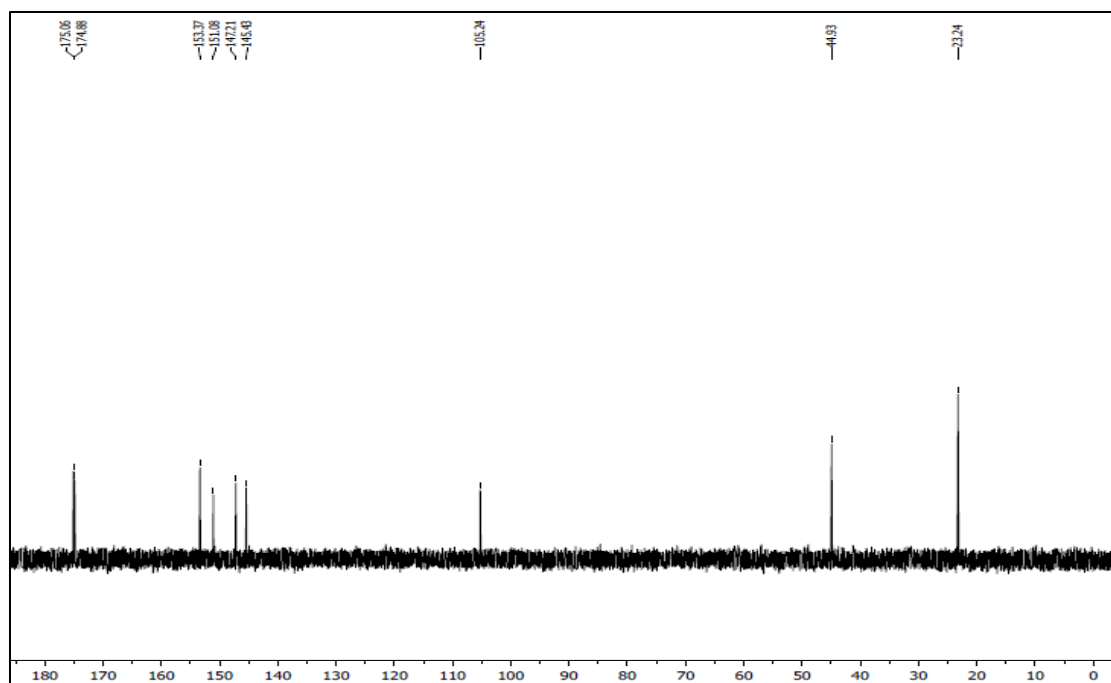
<sup>13</sup>C NMR of Compound **18**



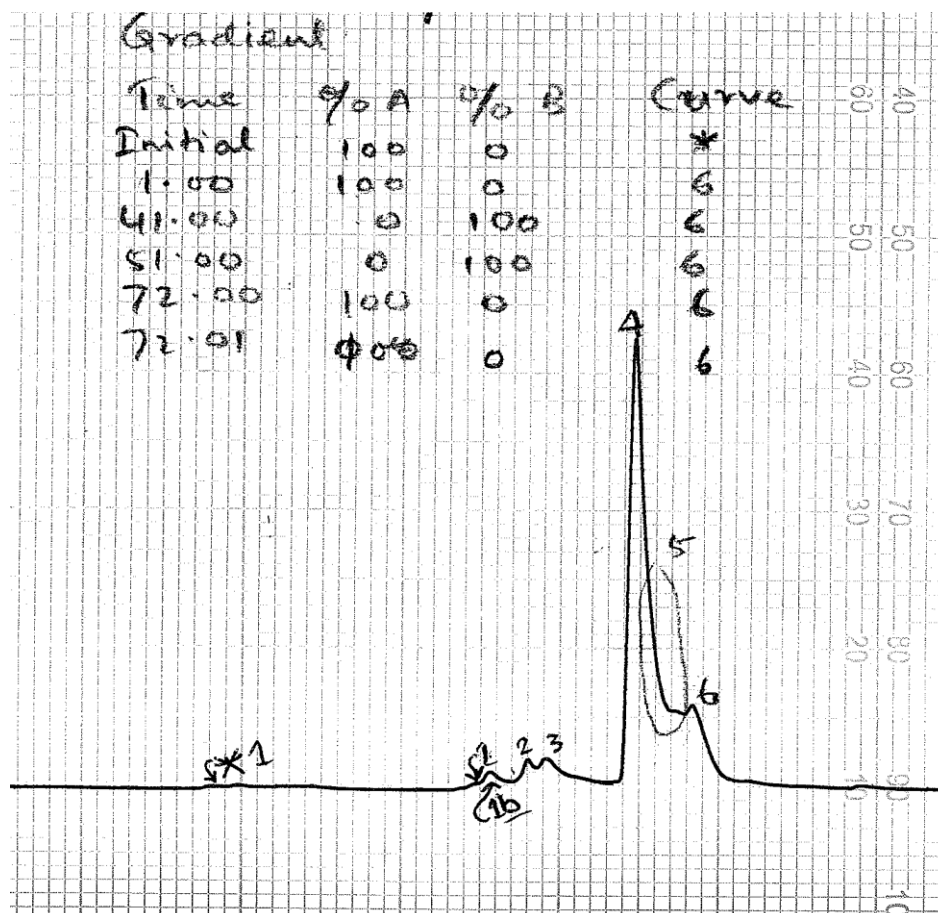
$^1\text{H}$  NMR of Compound **19**



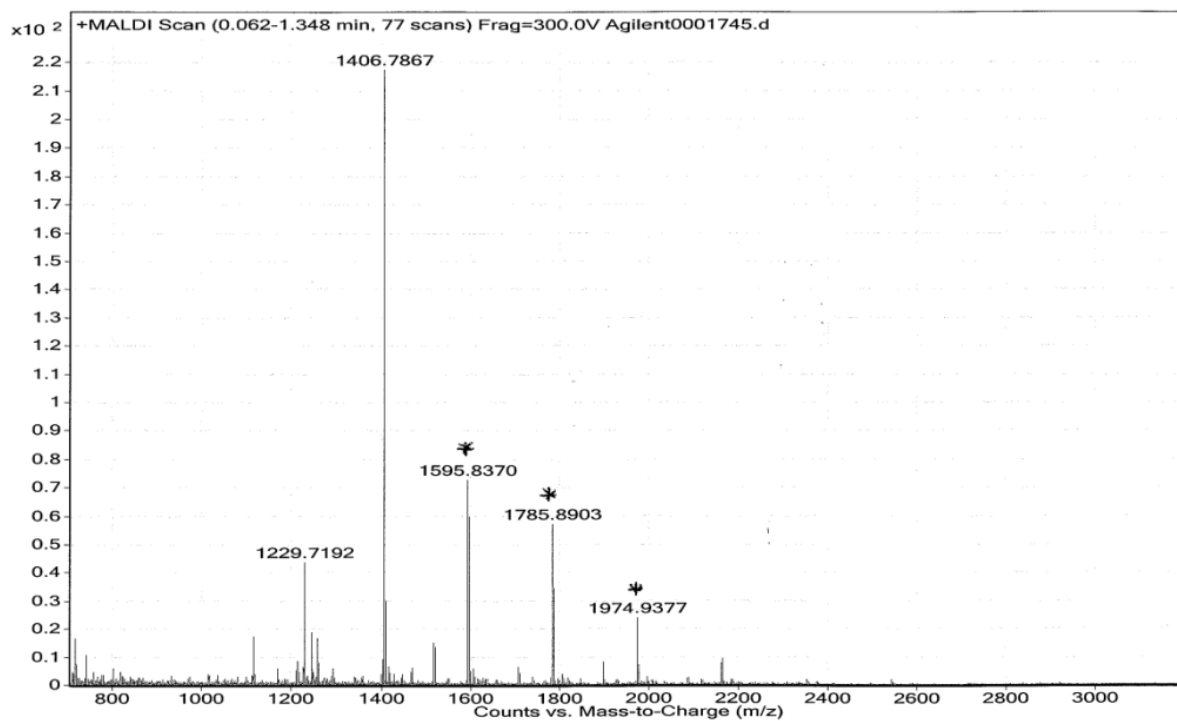
$^{13}\text{C}$  NMR of Compound **19**



# HPLC for the Model PNA sequence (Figure 4.12).



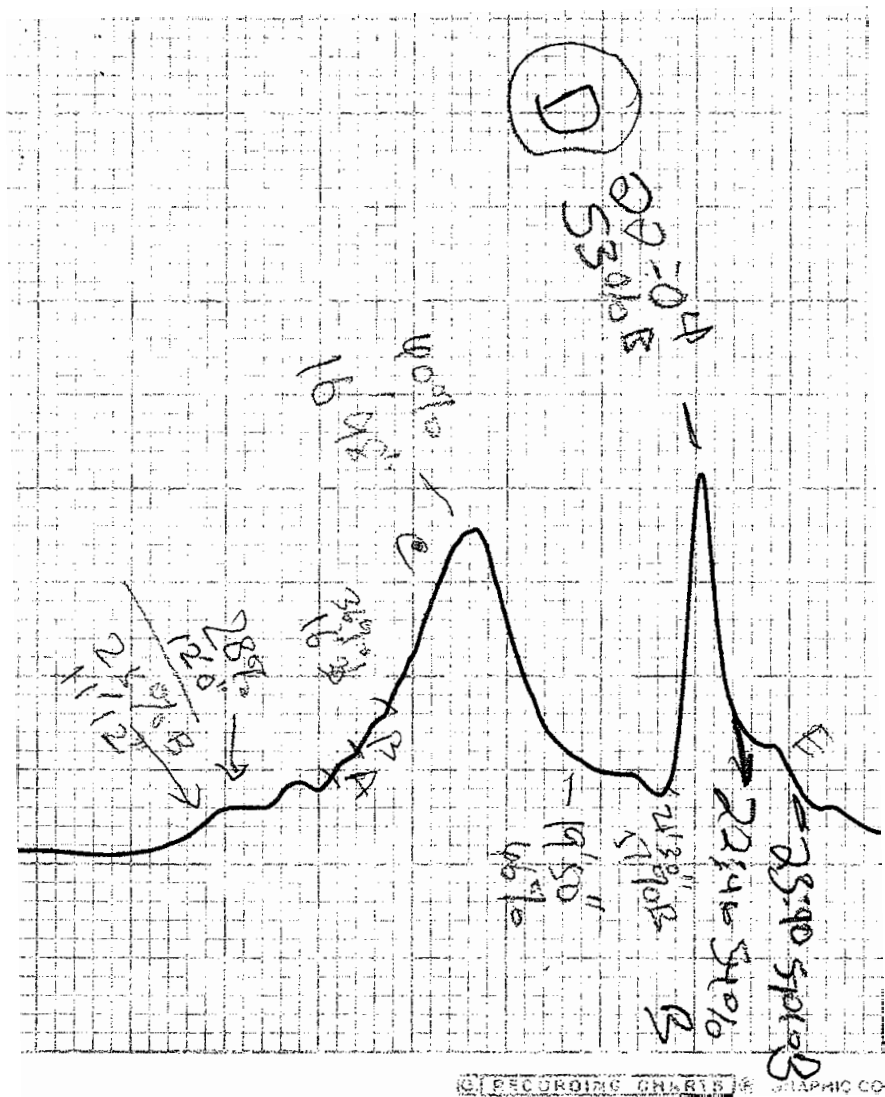
The region 4 of the HPLC trace contained the desired PNA sequence. Region 5 had some impurity. The MALDI-TOF mass spectrometry of the pure PNA is below.



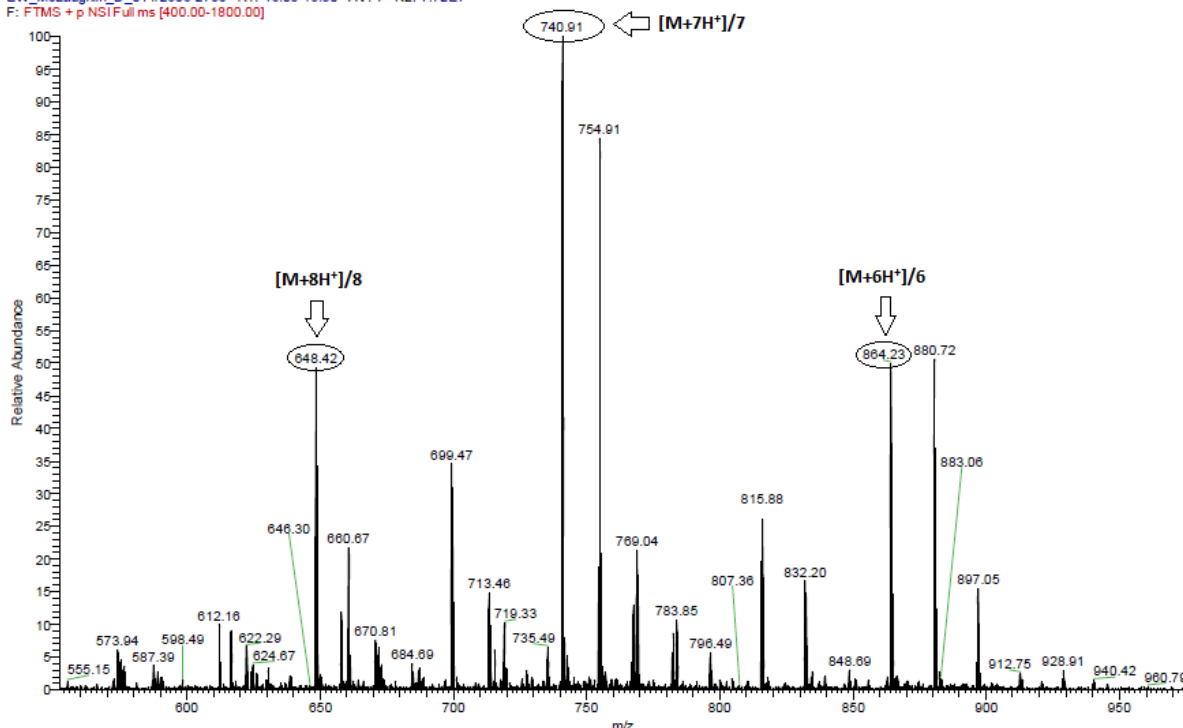
\* Matrix adduct peaks.



HPLC for the PNA sequence (Figure 4.11).



The region D of the HPLC trace contained the desired PNA sequence with impurity. The MALDI-TOF mass spectrometry of the pure PNA is below.



The circled masses corresponded to the multiply-charged PNA sequences. Unfortunately, the sample is not clean. We are trying to find a system to increase the purity on HPLC.

## References

- <sup>1</sup> [http://www.ornl.gov/sci/techresources/Human\\_Genome/home.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml)
- <sup>2</sup> Obika, S. *Chem. Pharm. Bull.* **2004**, 52(12), 1399—1404.
- <sup>3</sup> a) Plum, G. E.; Park, Y-W.; Singleton, S. F.; Dervan, P. B.; Breslauer, K. J. *Proc. Natl. Acad. Sci. USA* **1990**, 87, 9436-9440. b) Beal, P. A.; Dervan, P. B. *Science* **1991**, 251, 1360-1363.
- <sup>4</sup> a) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991**, 254, 1497-1500.  
 b) Egholm, M.; Buchardt, O.; Nielsen, P. E.; Berg, R. H. *J. Am. Chem. Soc.* **1992**, 114,

---

1895-1897. c) Egholm, M.; Buchardt, O.; Nielsen, P. E.; Berg, R. H. *J. Am. Chem. Soc.* **1992**, *114*, 9677-9678.

<sup>5</sup> a) Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023-2024. b) Felsenfeld, G.; Rich, A. *Biochim. Biophys. Acta* **1957**, *26*, 457-468 .

<sup>6</sup> a) Hoogsteen, K. *Acta Crystallogr.* **1959**, *12*, 822-823. b) Hoogsteen, K. *Acta Crystallogr.* **1963**, *16*, 907-916.

<sup>7</sup> Wang, G.; Levy, D. D.; Seidman, M. M.; Glazer, P. M. *Mol. Cell. Biol.* **1995**, *15*, 1759-1768.

<sup>8</sup> Soyfer, V. N.; Potaman, V. N. “*Triple-Helical Nucleic Acids*,” Springer-Verlag, New York, **1995**.

<sup>9</sup> a) Letai, A. G.; Palladino, M. A.; Fromm, E.; Rizzo, V.; Fresco, J. R. *Biochemistry* **1988**, *27*, 9108—9112. b) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* **1988**, *241*, 456-459. c) Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360-1363. c) Broitman, S. L.; Im, D. D.; Fresco, . R. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 5120-5124.

<sup>10</sup> Olivas, W. M.; Maher, L. Jr. *Nucleic Acid Res.* **1995**, *23*, 1936-1941.

<sup>11</sup> Cheng, A. J.; Van Dyke, M. W. *Nucleic Acid Res.* **1993**, *21*, 5630-5635.

<sup>12</sup> Faruqi, A. F.; Krawczyk, S. H.; Matteucci, M. D.; Glazer, P. M. *Nucleic Acids Res.* **1997**, *25*, 633—644.

<sup>13</sup> Branda, N.; Kurz, G.; Lehn, J. M. *Chem. Commun.* **1996**, 2443-2444.

<sup>14</sup> a) Nielsen, P. E.; Egholm, M.; Buchardt, O. *J. Mol. Recognit.* **1994**, *7*, 165-70. b) Jensen, K. K.; Orum, H.; Nielsen, P. E.; Norden, B. *Biochemistry* **1997**, *36*, 5072-7. c) Nielsen, P. E. & Egholm, M. *Bioorg. Med. Chem.* **2001**, *9*, 2429-34.

- 
- <sup>15</sup> Manning, G. S. *Biopolymers* **1983**, 22, 689-729.
- <sup>16</sup> Chen, D.; Meena; Sharma, S. K.; McLaughlin, L. W. *J. Am. Chem. Soc.*, **2004**, 126 (1), 70–71.
- <sup>17</sup> Chen, H.; Meena; McLaughlin, L. W. *J. Am. Chem. Soc.*, **2008**, 130 (40), 13190–13191.
- <sup>18</sup> Chen, C. H. B.; Sigman, D. S. *J. Am. Chem. Soc.* **1988**, 110, 6570-6572.
- <sup>19</sup> Montgomery, J. A.; Hewson, K. *J. Am. Chem. Soc.* **1960**, 82, 463-468.
- <sup>20</sup> Torii, T.; Yamashita, K.; Kojima, M.; Suzuki, Y.; Hijiya, T.; Izawa, K. *Nucleosides, Nucleotides and Nucleic Acids* **2006**, 25, 625-634.
- <sup>21</sup> Ram, S.; Spicer, L. D. *Syn. Comm.* **1987**, 17(4), 415-418.
- <sup>22</sup> a) Uhlmann, E.; Peyman, A.; Breipohl, G.; Will, D. W. *Angew. Chem., Int. Ed.* **1998**, 37, 2796–2823. b) Bialy, L.; Diaz-Mochon, J. J.; Specker, E.; Keinicke, L.; Bradley, M. *Tetrahedron* **2005**, 61, 8295–8305. c) Wojciechowski, F.; Hudson, R. H. *J. Org. Chem.* **2008**, 73, 3807–3816. d) Debaene, F.; Winssinger, N. *Org. Lett.* **2003**, 5, 4445–4447. e) Musumeci, D.; Roviello, G. N.; Valente, M.; Sapio, R.; Pedone, C.; Bucci, E. M. *Biopolymers* **2004**, 76, 535–542.
- <sup>23</sup> Husken, N.; Gasser, G.; Koster, S. D.; Metzler-Nolte, N. *Bioconjugate Chem.* **2009**, 20, 1578–1586.
- <sup>24</sup> Egholm, M.; Casale, R. A. *In Solid-phase Synthesis*; Kates, S. A., Albericio, F., Eds.; Dekker: New York, **2000**, 549–578.
- <sup>25</sup> Avitabile, C.; Moggio, L.; D’Andrea, L. D.; Pedone, C.; Romanelli, A. *Tetrahedron Letters* 51 (2010) 3716–3718
- <sup>26</sup> Adopted from *Applied Surface Science* **2006**, 252, 6935–6940

- 
- <sup>27</sup> Hansen, M. E.; Bentin, T.; Nielsen, P. E. *Nucleic Acid Research* **2009**, *37*(13), 4498-4507
- <sup>28</sup> Genadiy Kalayanova, G.; Jakšaa, S.; Scarciac, T.; Kobe, J. *Synthesis* **2004**, *12*, 2026-2034
- <sup>29</sup> Holmes, R. E.; Robins, R. K. *J. Am. Chem. Soc.* **1965**, 1772-1776
- <sup>30</sup> Bodepudi, V.; Shibutani, S.; Johnson, F. *Chem. Res. Toxicol.* 1992,5, 608-617.
- <sup>31</sup> Madre, M.; Panchenko, N.; Zhuk, R.; Geenevasen, J. A.; van den Burg, A; Koomen, G-  
J. *Synthesis* **1999**, *5*, 775-778.
- <sup>32</sup> Dunitz, J. D.; Taylor, R. *Chemistry* **1997**, *3*(1), 89-98.
- <sup>33</sup> Cobo, J.; Sánchez, A.; Nogueras, M. *Tetrahedron* **1998**, *54*, 5753-5762.